

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 19:14:53 ; Search time 185 Seconds
(without alignments)

3844.173 Million cell updates/sec

Title: US-09-522-753-5

Perfect score: 13215

Sequence: 1 MSGSTOLVAQWTRATEPRYP.....WDEPKPLLCQYETLSDSE 2517

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	13178.5	99.7	2518	3	AAB40574
2	12978	98.2	2507	6	ABU61812
3	10796.5	81.7	2462	6	ABU61813
4	7866	59.5	1495	2	AAW18226
5	4611	34.9	876	7	ADC35088
6	4187	31.7	2453	3	AAW12454
7	4043	30.6	2343	3	AAW12453
8	1113	8.4	216	6	ABU70857
9	1060	8.0	3502	4	ABBS8382
10	960	7.3	184	6	ABU70980
11	774.5	5.9	224	6	ABU70979
12	718	5.4	619	2	AAW99738
13	569.5	4.3	3371	6	ABO07211
14	569.5	4.3	3664	6	ABR47592
15	569.5	4.3	3664	6	ABO53027
16	565.5	4.3	3266	3	AAW42491
17	552.5	4.2	2406	6	ABW82733
18	550	4.2	1963	4	ABW62819
19	539.5	4.1	8991	6	ABU08487
20	533	4.0	3080	4	ABW64877
21	523	4.0	2639	4	ABG15016
22	522.5	4.0	2133	4	AAU33195
23	519	3.9	3536	4	ABW65480
24	518	3.9	2665	4	AAW14533
25	518	3.9	2665	4	ABB33490

26	518	3.9	2665	4	AAW26950
27	518	3.9	2665	4	ABB28314
28	518	3.9	2665	4	ABW18950
29	518	3.9	2665	4	AAW66665
30	518	3.9	2665	4	AAW54270
31	518	3.9	2665	4	ABG48336
32	518	3.9	2665	4	AAW02259
33	518	3.9	2665	5	ABG36319
34	512.5	3.9	2703	4	ABW63299
35	512.5	3.9	2703	5	ABG70019
36	494.5	3.7	2724	4	ABG20119
37	485.5	3.7	2803	5	ABW08161
38	477.5	3.6	3118	4	AAW50362
39	473	3.6	3201	6	ABW62899
40	469.5	3.6	2701	6	ABR92087
41	469.5	3.6	2819	4	ABW35408
42	466.5	3.5	5560	4	ABW71160
43	458.5	3.5	3111	4	ABW60327
44	458.5	3.5	5533	4	ABW65772
45	458	3.5	1708	5	ABW69375
46	456	3.5	2971	3	AAW41231
47	452.5	3.4	1566	6	ABR59716
48	452	3.4	1612	4	ABW59410
49	452	3.4	2971	7	ADC35084
50	452	3.4	2972	4	AAW50363
51	451.5	3.4	2193	6	ABR42219
52	450.5	3.4	2735	6	ABR41356
53	450.5	3.4	2768	4	ABW68397
54	441	3.3	2759	4	ABW71148
55	440	3.3	2897	4	ABW58514
56	436	3.3	4019	7	AAE13839
57	436	3.3	4019	7	ADD66733
58	436	3.3	4019	7	ADW87987
59	436	3.3	4025	5	ABW69736
60	430	3.3	2951	4	ABW60291
61	429.5	3.3	2429	5	AAE21713
62	429	3.2	2091	3	AAW12000
63	428	3.2	1299	4	AAW24322
64	428	3.2	1404	2	AAW26049
65	428	3.2	1404	4	AAW29773
66	428	3.2	1404	4	AAW60568
67	428	3.2	1415	4	AAU32262
68	428	3.2	2063	4	AAW78710
69	428	3.2	2063	4	AAW40064
70	428	3.2	2063	4	ABG17147
71	428	3.2	2063	6	ABG74436
72	425.5	3.2	2545	5	ABW98406
73	424.5	3.2	2429	7	AAW78989
74	424.5	3.2	2523	4	AAU03503
75	422	3.2	2087	4	AAW31516
76	422	3.2	2157	7	ABO44406
77	422	3.2	2217	7	ABO44390
78	422	3.2	2294	7	ABO44405
79	422	3.2	2354	7	ABO44389
80	418	3.2	1828	4	ABW50164
81	417.5	3.2	1938	6	ABW76681
82	417	3.2	2005	6	ABG74435
83	416.5	3.2	1791	6	ABO52986
84	415.5	3.1	1266	6	ABW52986
85	415.5	3.1	1638	7	ADC31236
86	414	3.1	2092	7	ABW78990
87	414	3.1	2161	4	AAW78959
88	414	3.1	2189	4	AAW79943
89	413	3.1	4873	6	ABO14747
90	410	3.1	2468	6	ABR64281
91	410	3.1	2468	7	ADW62723
92	410	3.1	2468	7	ADW62719
93	410	3.1	2468	7	ADW62727
94	410	3.1	2468	7	ADW62715
95	410	3.1	2519	4	ABG16636
96	408	3.1	2185	7	ABO44398
97	408	3.1	2245	7	ABO44382
98	408	3.1	2322	7	ABO44397

99 408 3.1 2382 7 ABO44381 Abo44381 Novel hum
100 406.5 3.1 1565 4 AAM41059 Aam41059 Human pol

ALIGNMENTS

RESULT 1
AAB40574
ID AAB40574 standard; protein; 2518 AA.
XX AAB40574;
AC AAB40574;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antichyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
PI
XX WPI; 2000-602362/57.
DR N-PSDB; AAC74783.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 784-790; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antineumatic; antichyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 2518 AA;
Query Match 99.7%; Score 13178.5; DB 3; Length 2518;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2512; Conservative 1; Mismatches 4; Indels 1; Gaps 1;
QY 1 MSGSTQLVAQTWRATEPRYPHSLSYVQIARTHTDVGLLEYQHHSRDYASHLSPGSIQ 60
DB 1 MSGSTQLVAQTWRATEPRYPHSLSYVQIARTHTDVGLLEYQHHSRDYASHLSPGSIQ 60
QY 61 PORRPSLLSEFQPGNERSQELHLRPESHSHYLPELGKSEMEFIESKRPLELLPDLRLP 120
DB 61 PORRPSLLSEFQPGNERSQELHLRPESHSHYLPELGKSEMEFIESKRPLELLPDLRLP 120
QY 121 SPLLATGQAGSEDLTKDRSLTGKLEPVSPPPHTDPELELVPPRLSKEELIQNMDRVD 180
DB 121 SPLLATGQAGSEDLTKDRSLTGKLEPVSPPPHTDPELELVPPRLSKEELIQNMDRVD 180
QY 181 REITWVEQIISKKKQOQLEEEAAKPPPEKPVSPPIESKHSRLVQIYDENRKAAB 240
DB 181 REITWVEQIISKKKQOQLEEEAAKPPPEKPVSPPIESKHSRLVQIYDENRKAAB 240
QY 241 AHRILEGLGPQVELYLNQPSDTROYHENIKINOAMRKKLILYFKRNHARQWKQFCQ 300
DB 241 AHRILEGLGPQVELYLNQPSDTROYHENIKINOAMRKKLILYFKRNHARQWKQFCQ 300
QY 301 RYDQLEALEKKKVERIENPRRAKESKVREYVEKQFPEIRKORELOERMQRVGGSG 360
DB 301 RYDQLEALEKKKVERIENPRRAKESKVREYVEKQFPEIRKORELOERMQRVGGSG 360
QY 361 LSMSAARSEHEVSEIIDGLSEQENLEKQMRQLAVIPPMYLDADQQRIKFINNGLMADPM 420
DB 361 LSMSAARSEHEVSEIIDGLSEQENLEKQMRQLAVIPPMYLDADQQRIKFINNGLMADPM 420
QY 421 KYKDRQVNMWMSQEKETFREKFMQHPKNGFLIASFLERKTVAECVLYYITKKENYK 480
DB 421 KYKDRQVNMWMSQEKETFREKFMQHPKNGFLIASFLERKTVAECVLYYITKKENYK 480
QY 481 SLVRSYRRRGKS-QQQ 539
DB 481 SLVRSYRRRGKS-QQQ 540
QY 540 NDKEDLLKEKTDGSDNDEKEAVASKGRKTANSQGRKGRITRSMANEANSEEAITPQ 599
DB 541 NDKEDLLKEKTDGSDNDEKEAVASKGRKTANSQGRKGRITRSMANEANSEEAITPQ 600
QY 600 QSAELASMLNNESSRWTEEMETAKGILLEHGRNWSAIARMVGSKTVSCKKPFYFNKXR 659
DB 601 QSAELASMLNNESSRWTEEMETAKGILLEHGRNWSAIARMVGSKTVSCKKPFYFNKXR 660
QY 660 QNLDELIOQHKLMEKERNARRKKKAPAAASEEAAPPVVEDEMEASGVSNEEMVVE 719
DB 661 QNLDELIOQHKLMEKERNARRKKKAPAAASEEAAPPVVEDEMEASGVSNEEMVVE 720
QY 720 EAEALHASGNEVPRGECGPGATVNNSSDTPESIPSPHTEAAKDTGQNGPKPPATLGADGPP 779
DB 721 EAEALHASGNEVPRGECGPGATVNNSSDTPESIPSPHTEAAKDTGQNGPKPPATLGADGPP 780
QY 780 PGPPPTPRTSRAPTEPTASATGATPPAPPSPSAPPVVPVPEKEEETAAAPVVE 839
DB 781 PGPPPTPRTSRAPTEPTASATGATPPAPPSPSAPPVVPVPEKEEETAAAPVVE 840
QY 840 GGEQKPPAAEELAVDTGKAEFPVKSECTEAESEGPAGKGDAAEAATAEGALKAEKKEGG 899
DB 841 GGEQKPPAAEELAVDTGKAEFPVKSECTEAESEGPAGKGDAAEAATAEGALKAEKKEGG 900
QY 900 SGRATTAKSSGAPQSDSSATCSADEVDBAEGDKNRLSLPRPSLLTPTGDPANASPOK 959


```

QY 1733 DLQVPHLPVLVPTTCTATAMDRLAYLPTAQPFSSRRSSPLSPGGGTHLTKTPTTS 1792
Db 1723 DLQVPHLPVLVPTTCTATAMDRLAYLPTAQPFSSRRSSPLSPGGGTHLTKTPTTS 1782
QY 1793 SSERDRDRDRDREREKSLTSTTVEHAPIMRPGTEQSSGSSGSGGSSSRPA 1852
Db 1783 SSERDRDRDRDREREKSLTSTTVEHAPIMRPGTEQSSGSSGSGGSSSRPA 1842
QY 1853 SHSHAHQHSPISPRTQDALQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSPVRPAAT 1912
Db 1843 SHSHAHQHSPISPRTQDALQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSPVRPAAT 1902
QY 1913 FPPATHCPGLGTLGVYPTLMERVLVLPKEAPRVARPERPADTGHAFAPKPARSGLEPA 1972
Db 1903 FPPATHCPGLGTLGVYPTLMERVLVLPKEAPRVARPERPADTGHAFAPKPARSGLEPA 1962
QY 1973 SSPSKGSEPRPLVPVSGHATIAARTPAKNLAPHASDPDPAPASADPHREKTSQKPS 2032
Db 1963 SSPSKGSEPRPLVPVSGHATIAARTPAKNLAPHASDPDPAPASADPHREKTSQKPS 2022
QY 2033 IQELELRSLGYHGSSYSPEGEVPSVSPSLTHDKGLPKHLEELDKSHLEGELRPKQPG 2092
Db 2023 IQELELRSLGYHGSSYSPEGEVPSVSPSLTHDKGLPKHLEELDKSHLEGELRPKQPG 2082
QY 2093 PVKLGEEAHLPHLRPLPESQSSSPLLQTPAGVKGHQVRVTLAQHISEVITQDTRHP 2152
Db 2083 PVKLGEEAHLPHLRPLPESQSSSPLLQTPAGVKGHQVRVTLAQHISEVITQDTRHP 2142
QY 2153 QQLSAPLAPLYSPFGASCPLDLRRPPSDLYLPPDPHGAPARGSPHSEGGKSPENKT 2212
Db 2143 QQLSAPLAPLYSPFGASCPLDLRRPPSDLYLPPDPHGAPARGSPHSEGGKSPENKT 2202
QY 2213 SVLGGGEDGTEPVSPPEGTEPGHSSAVYPLLYRDGEQTEPSRMGSKPGNTSQPPAF 2272
Db 2203 SVLGGGEDGTEPVSPPEGTEPGHSSAVYPLLYRDGEQTEPSRMGSKPGNTSQPPAF 2262
QY 2273 SKLTESNANVSKKQKQKINKNLTHNRNEPEYNISQFTEIFNMPAITGTGLMYRSQAV 2332
Db 2263 SKLTESNANVSKKQKQKINKNLTHNRNEPEYNISQFTEIFNMPAITGTGLMYRSQAV 2322
QY 2333 QEHASTNMGLEALIRKALMGKYDQWESPPLSANFNPLNASASLPAAMPITAADGRSDH 2392
Db 2323 QEHASTNMGLEALIRKALMGKYDQWESPPLSANFNPLNASASLPAAMPITAADGRSDH 2382
QY 2393 TLTSFGGGKAKVSGRPSRKAKSPAGLASGRDPPSVSVHSEGCNRRTPLTNRWED 2452
Db 2383 TLTSFGGGKAKVSGRPSRKAKSPAGLASGRDPPSVSVHSEGCNRRTPLTNRWED 2442
QY 2453 RPSSAGSTPPFYNPLIMRLQAGVWASPPPPGLPAGSGPLAGPHAWDEEPKPLLCQYET 2512
Db 2443 RPSSAGSTPPFYNPLIMRLQAGVWASPPPPGLPAGSGPLAGPHAWDEEPKPLLCQYET 2502
QY 2513 LSDSE 2517
Db 2503 LSDSE 2507

RESULT 3
ID ABU61813
XX AC ABU61813;
XX AC ABU61813;
DT 12-AUG-2003 (first entry)
DE Mouse nuclear receptor corepressor SMRte.
XX Mouse; SMRte; nuclear receptor corepressor; gene therapy; tissue typing;
KW Mouse; SMRte; nuclear receptor corepressor; gene therapy; tissue typing;
XX cancer.
OS Mus musculus.
XX

```

```

FH Key Location/Qualifiers
FT Domain 166..430
FT /label= SNC_domain
FT Region 166..207
FT /label= Amphipathic helix
FT /notes= Residues 172, 180, 187, 194 and 201 make up the
FT hydrophobic heptad repeat"
FT Domain 431..487
FT /label= SANT_A_domain
FT Domain 610..666
FT /label= SANT_B_domain
XX
XX US2003027137-A1.
XX
XX 06-FEB-2003.
XX
XX 27-MAR-2001; 2001US-00819104.
XX
XX 29-MAR-2000; 2000US-0193138P.
XX
XX (CHEN/) CHEN J D.
XX
XX Chen JD;
XX
XX WPI; 2003-466139/44.
XX
XX N-PSDB; ACA62451, ACA62452.
XX
XX New SMRte proteins and nucleic acids, useful in gene therapy, predictive
XX medicine, therapeutic or prophylactic treatment, chromosome mapping,
XX tissue typing and in forensic biology.
XX
XX Claim 9; Page 65-71; 90pp; English.
XX
XX The invention relates to an isolated SMRte nucleic acid molecule. The
XX nucleic acids are useful in gene therapy, as hybridisation probes for
XX identifying SMRte-encoding nucleic acid molecules and as primers for
XX amplifying of SMRte nucleic acid molecules. The polypeptides are useful
XX as immunogens to raise anti-SMRte antibodies. The SMRte molecules are
XX useful as targets for discovering and developing modulating agents to
XX regulate a variety of cellular processes, in screening assays, in
XX predictive medicine, in therapeutic or prophylactic treatment, in
XX chromosome mapping, tissue typing and in forensic identification of a
XX biological sample. Modulators of SMRte are useful for treating or
XX preventing a condition associated with aberrant SMRte protein or nucleic
XX acid expression or activity, such as cancer. The present sequence
XX represents the amino acid sequence of the mouse nuclear receptor
XX corepressor SMRte
XX
XX Sequence 2462 AA;

```

```

Query Match 81.7%; Score 10796.5; DB 6; Length 2462;
Best Local Similarity 83.4%; Pred. No. 0;
Matches 2126; Conservative 94; Mismatches 210; Indels 119; Gaps 36;

```

```

QY 1 MSGSTQLVAQTWRATEPRYPHLSYVPQIARTHTDVGLEVOHHSRDYASHLSGSIQ 60
Db 1 MSGSTQVQAQTWRAEPRYPHLSYVPQIARSHTDVGLEVOHHSRDYASHLSGSIQ 60
QY 61 PQRPRPSLLSEFQPGNERSQELHLRPESHSLYLPDLCKSEMEFIESKRPRLLELPDLRP 120
Db 61 PQRPRPSLLSEFQPGNERSQELHLRPESHSLYLPDLCKSEMEFIESKRPRLLELPDLRP 120
QY 121 SPLLATGQPGAGSEDLTKDRSLTKLEPVPSPPHPTDPELELVPPRLSKBELQN-MDRV 179
Db 121 SPLLATGQPGAGSEDLTKDRSLTKLEPVPSPPHPTDPELELVPPRLSKBELQN-MDRV 180
QY 180 DREITWVEQIISKLKKKQQLSEEAAPPEPEKPVSPPPPIESKHSRLVQIYYDENRKA 239
Db 181 DREITWVEQIISKLKKKQQLSEEAAPPEPEKPVSPPPPIESKHSRLVQIYYDENRKA 240
QY 240 AAHRILEGLGQVPELPLYNQPSDTROYHENIKINQAMRKKLILYFKRRNHARKQWKPC 299
Db 241 AAHRILEGLGQVPELPLYNQPSDTROYHENIKINQAMRKKLILYFKRRNHARKQWKPC 300

```


||||| 2374 SVSSVHSEGCNRTPLTRNWRDRESSAGSTFPYPLIMRLQAGVMASPPPPGLAAGS 2433
 QY 2489 GPLAGPHHAWDEPKPLLCQSYETLSDSE 2517
 Db 2434 GPLAGPHHAWDEPKPLLCQSYETLSDSE 2462

RESULT 4
 AAW18226
 ID AAW18226 standard; protein; 1495 AA.
 XX AAW18226;
 AC AAW18226;
 XX 24-SEP-1997 (first entry)
 DE Transcriptional co-repressor SMRT.
 XX Silencing mediator for retinoic acid and thyroid hormone receptor; SMRT;
 KW transcriptional co-repressor.
 XX Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Region 1. .160
 FT /label= N-terminal region
 FT /note= "proline-rich domain"
 FT Region 773. .790
 FT /label= ERDR_region
 FT Region 812. .827
 FT /label= SG_region
 FT Region 1061. .1132
 FT /label= glutamine-rich region
 FT Region 1201. .1495
 FT /label= C-terminal_region
 FT Peptide 1330. .1376
 FT /note= "alternatively spliced insert not present in the
 FT original two-hybrid clone"
 XX WO9709418-A1.
 PN 13-MAR-1997.
 PD 24-JUL-1996; 96WO-US012371.
 PF 01-SEP-1995; 95US-00522726.
 PR (SALK) SALK INST BIOLOGICAL STUDIES.
 XX Evans RM, Chen JD;
 PI WPI; 1997-192894/17.
 DR New co-suppressor of steroid-thyroid hormone receptor activity - also
 XX methods for identifying compounds that relieve its suppressant effect
 PT and/or activate receptors.
 XX Claim 2; Page 40-45; 71pp; English.
 PS A novel receptor interacting factor (AAW18226) is designated SMRT, i.e.
 CC silencing mediator (co-repressor) for retinoic acid receptor (RAR) and
 CC thyroid hormone receptor (TR). Its association with RAR and TR both in
 CC solution and on DNA response elements is destabilised by ligand. The
 CC interaction of SMRT with mutant receptors correlates with the
 CC transcriptional silencing activities of receptors. In vivo, SMRT
 CC functions as a potent co-repressor. A GAL4 DNA binding domain fusion of
 CC SMRT behaves as a frank repressor of a GAL4-dependent reporter. These
 CC data identify a novel class of cofactor which is believed to represent an
 CC important mediator of hormone action. Full-length cDNA for SMRT has been
 CC isolated from a HeLa library in a two-hybrid screen using a GAL4 DNA
 CC binding domain/RXR fusion protein as bait
 XX Sequence 1495 AA;

Query Match 59.5%; Score 7866; DB 2; Length 1495;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1486; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1031 DKEAFAAEAQKLPDPCWTSGLPPVPREVIKASPHAPDPSAFYAPRGHPLPLGLHD 1090
 Db 9 DKEAFAAEAQKLPDPCWTSGLPPVPREVIKASPHAPDPSAFYAPRGHPLPLGLHD 68
 QY 1091 TARPVLPRPPTISNPPPLISSAKHPSVLERQICAIISQGMVOLHVYPSHAKAPVGPVTM 1150
 Db 69 TARPVLPRPPTISNPPPLISSAKHPSVLERQICAIISQGMVOLHVYPSHAKAPVGPVTM 128
 QY 1151 GLPLPMDPKLAPFSGVKQELSPRGQAGPPESLGVPTAQEASVLRGTALGSPVGSITK 1210
 Db 129 GLPLPMDPKLAPFSGVKQELSPRGQAGPPESLGVPTAQEASVLRGTALGSPVGSITK 188
 QY 1211 GIPSTRVPDSAITYRGSIHTGTPADVLYKGTITRIIGDPSRLDRGREDSLPKGHVY 1270
 Db 189 GIPSTRVPDSAITYRGSIHTGTPADVLYKGTITRIIGDPSRLDRGREDSLPKGHVY 248
 QY 1271 EGKKGHVLSYEGGMSVTQCSKEDGRSSGPPHETAAKRTYDMMEGRVGRAISSASIEGL 1330
 Db 249 EGKKGHVLSYEGGMSVTQCSKEDGRSSGPPHETAAKRTYDMMEGRVGRAISSASIEGL 308
 QY 1331 MGRAIPPERHSPHLLKEQHHRIGSITQGIIPRSYVEAQEDYLREAKLLKREGTPPPPPS 1390
 Db 309 MGRAIPPERHSPHLLKEQHHRIGSITQGIIPRSYVEAQEDYLREAKLLKREGTPPPPPS 368
 QY 1391 RDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPREELRHTPELPAPRPLKEGS 1450
 Db 369 RDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPREELRHTPELPAPRPLKEGS 428
 QY 1451 ITQGTPLKYDTGASTTGSKKHDVRSLSIGSPGRTPFPVHPLDVMDARALACRYESLSKS 1510
 Db 429 ITQGTPLKYDTGASTTGSKKHDVRSLSIGSPGRTPFPVHPLDVMDARALACRYESLSKS 488
 QY 1511 RFGTASSGGSIARGAPVIVPELGPQSPLYVEDHGAPFAGHLPGRGSPVTMEPTPRLQ 1570
 Db 489 RFGTASSGGSIARGAPVIVPELGPQSPLYVEDHGAPFAGHLPGRGSPVTMEPTPRLQ 548
 QY 1571 EGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSHIPL 1630
 Db 549 EGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSHIPL 608
 QY 1631 AFDPTSIIPRGIPLDRAAAAYLPRHLAPNPTYPHLYPPYLIRGYDPTAALENROTIIINDYI 1690
 Db 609 AFDPTSIIPRGIPLDRAAAAYLPRHLAPNPTYPHLYPPYLIRGYDPTAALENROTIIINDYI 668
 QY 1691 TSQQMHNTATAMAQADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPTPGT 1750
 Db 669 TSQQMHNTATAMAQADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPTPGT 728
 QY 1751 PATAMDRLAYLPTAPQPFSSRHSSPLSPGGPHLTPTTTSSSERERDRDRDRDR 1810
 Db 729 PATAMDRLAYLPTAPQPFSSRHSSPLSPGGPHLTPTTTSSSERERDRDRDRDR 788
 QY 1811 EKSILTSTTTVSHAPIWRPCTQSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 1870
 Db 789 EKSILTSTTTVSHAPIWRPCTQSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 848
 QY 1871 LQORPSVLHNTGMKGIIITAVEPSKPTVLARSTSTSSPVRPAATPPATHCPGLGTLDGVYP 1930
 Db 849 LQORPSVLHNTGMKGIIITAVEPSKPTVLARSTSTSSPVRPAATPPATHCPGLGTLDGVYP 908
 QY 1931 TLMEPVLLPKEAPRVARPERPRADTGHAFIAPKPARSGLEPASPSSKSGSEPRPLVPVSG 1990
 Db 909 TLMEPVLLPKEAPRVARPERPRADTGHAFIAPKPARSGLEPASPSSKSGSEPRPLVPVSG 968
 QY 1991 HATIARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQEELSLGLVHGSSYSP 2050
 Db 969 HATIARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQEELSLGLVHGSSYSP 1028

Db 855 HAWDEPKLLCSQYETLSDSE 876

RESULT 6
AAB12454
ID AAB12454 standard; protein; 2453 AA.
XX
AC AAB12454;
DT 24-OCT-2000 (first entry)
XX
DE HNRCR protein sequence.
XX
KW Human; HNRCR; nuclear receptor coreceptor.
XX
OS Unidentified.
XX
PN CN1250094-A.
PD 12-APR-2000.
PF 06-OCT-1998; 98CN-00120919.
XX
PR 06-OCT-1998; 98CN-00120919.
XX
PA (XINH-) XINHANGPU FUDAN GENE ENG CO LTD SHANGHA.
XX
PI Yu L, Tu Q, Zhao Y;
XX
DR WPI; 2000-400830/35.
DR N-PSDB; AAA60630.
XX
PT Preparation of new human kernon acceptor co-repressor coding series and the polypeptide.
XX
PS Example 2; Fig 2; 58pp; Chinese.
XX
CC The present invention describes a human homologue of nuclear receptor coreceptor (HNRCR). The present sequence represents an HNRCR protein sequence used in comparison with the human HNRCR
XX
SQ Sequence 2453 AA;

Query Match 31.7%; Score 4187; DB 3; Length 2453;
Best Local Similarity 41.0%; Pred. No. 2.1e-226;
Matches 1103; Conservative 348; Mismatches 796; Indels 444; Gaps 106;

QY 16 EPRYPHSLSYVQIARTHTDVLLEYQ--HHRDYASHLSGSIITQPQRRLPSILLSEFQ 73
DB 17 QSRYPHSHVQYTPFSARHQOEFAVPDYRSHLEVVSQASQLLOQQQQQLRRRPSILLSEF 76
QY 74 PGNERSQELHLRPESHLSYLPCLGKSEMEFTESKRPLELLDPDLLR-----PSPLLATGQ 128
DB 77 PGSDRQOE--RRGYYQFHFGPSVDHSLSKRPLEQVSDSHFORISAAVLPVHT-L 133
QY 129 PAG---SEDITKDRSLTGKLE-PVSPSPPTDPELVPRLSKKEELIQNMDRVDREIT 184
DB 134 PEGLRSSANAKDPAGVKGHEAPSSPLSGQPCGDDQNAPSKLSKEELIQSDMDRVDREIA 193
QY 185 MVEQQTSLKKKKQQLLEEEAAKPEPEKPVSPPIESKHSLSVQIYIDENRKAHAHRI 244
DB 194 KVEQQILKLLKKQQQLLEEEAAKPEPEKPVSPPIESKHSLSVQIYIDENRKAHAHRI 253
QY 245 LEGLGPVELPLYNQPSDTQYHENIKINQAMKLLIYFKRNHARKWKQKFCORYDQ 304
DB 254 FEGLGKPKVLPYNQPSDTQYHENIKINQAMKLLIYFKRNHARKWKQKFCORYDQ 313
QY 305 LMEALEKQVRIENPNRRRAKESKREYIEKQPEIRKQRELOERMQSRVQSGLSMS 364
DB 314 LMEAEKQVRIENPNRRRAKESKREYIEKQPEIRKQRELOERMQSRVQSGLSMS 372
QY 365 AARSEHEVSEIIDGLSEQENLEKQMLAVIPMLYDADQQRITKFINMGLMADPMKVYK 424

Db 373 IARSEHEISEIIDGLSEQENNEKQMLQSVIPPMFDBAQRVRVKFINMGLMEDPMKVYK 432
QY 425 DRQVMNMWSEQEKETPREKFMQHPKFNGLIASFLERKTVAECVLYLYLTKNENYKSLVR 484
Db 433 DRQFMNVWTDHEKEIFKDKFIQHPKFNGLIASFLERKSVPCDVLVLYLTKNENYKALVR 492
QY 485 RSY-RRRGKSQQQQQQQQQQQQQQQPMRPSQBEKDEKEKEKEAEK--EEREPEVEND 541
Db 493 RNYKRGGRNQIARPSQBEKVEEK-----EEDKAETKEKEKEKDEKDEKDEKDEK 544
QY 542 KEDLLKEKTDGSDGNDSEKAVASKRKTANSQGRKGRITRSMANE--ANSEBAITP 598
Db 545 KETTKKORTAETAEPEERQVTPRKTANSQGRKGRITRSMANSEAAAAAATAATE 604
QY 599 QOSAEI-----ASMELNESSRWTEEMETAKGLLEHGRNWSAIAARMVSGSKTVSQCKNF 652
Db 605 EPPPLPPPEPISTEPVETSRWTEEMEVAKGLVEHGRNMAAIAKMGVTKSEAQCKNF 664
QY 653 YFNYKGRQNLDELQOHKLMKERNARRKKAPAAASEEAAFPVPVDEEHEASVSG 712
Db 665 YFNYKGRHNLNLLQOHKQKASRKPREERDVSCESVASTVSA---QEDEDIAS--- 716
QY 713 NEEEMVEEAEALHASGNEVPRGCSGPAITVNSSDTESIPSPH-TEAAK---DTGONGPK 768
Db 717 NEEENPEDSEG-----AENSDDTESAPSPFVAAKSSSESENA-- 756
QY 769 PPATLGADGPPPGPPTPRRTSRAPTEPTASATGATPPPPAPPSPSPAPPVVPKEKE 828
Db 757 -----ASRGNTPEVALEATDPAFCASP--SSAVPTTKPAERES 794
QY 829 EE-----TAAAPPVEGEEQ-----KPPAAELAVDTGKAEBPVKSECTEAE 871
Db 795 VEAQVTDASAETAEPMDVDHECGAEGSVLDPAPATK--ADSVDPQMOV-PENTASKG 851
QY 872 EGPAGKDAEAEATAEGALKAEKGGSGRATTAKSSGAPO--DSDSATSCADSDVDE 928
Db 852 EGDAKERDLEST-----SEKTEARDEDVVVAQIERPEPQSDDDSDSATSCADSDVDE 902
QY 929 AEGGDKNRL--LSRPSLLTPTGDPANASQPKP--LDLKLQKQRAAAIIPPI----- 976
Db 903 GE-PERQVFPMDAKPSLLTPGSIIL-SPPIKPNLLDLPLQHQRAAVIPPMVSCITCNI 960
QY 977 -----QVTKVHEPPREDAAPTKAPPAPPNPQNLQPEDSDAPQQGSSPRGK 1022
Db 961 PIGTPVSGYALYQRIKAMHESALLE-----EQRQEQVDLECRSSTSPCST 1008
QY 1023 SRSPAPDKEAFAAAQKLPGDPCWTSLGPPVPREVIVKASPHAPDPSAFSYPAGH 1082
Db 1009 SKSP-----NRE-----W-----EVLQAPH-----QVITNLEGV 1034
QY 1083 PLPLGLHDTARVLPRLPPTISNPPPLISSAKHPSVLERQIGALSQ-----MSVQLHVPY 1137
Db 1035 RLP-----TTRTPRPPPLIPSKTIVASEK-PSFI--MGSGISQGTGTYLSSHNAQYP 1086
QY 1138 SEHAKAPGVVWGLPLPMDPKLAPFSGVKQQLSPRQAGPPELGVPTAEASVLRG 1197
Db 1087 QEAPKPSVGSISLGLPRQOESTKAAPLTVIKQEFSPRSONSQPEGLLV-RAQHEGVVRG 1145
QY 1198 TALGSPVGSITKIGIPSTRVPSDSAITYRGSITHTGPA-----DVLKGTITRI-IG 1248
Db 1146 TA-GAVQESITGTTPASKISVETISSLRGSIQTGTPALPQAGIPTALVKGVPSRMPLE 1204
QY 1249 EDSPSRLDRGSDSLPKGHVIEGKKGHVLSYEGGMSVTQCSKEDGRSSGPPHETAAPK 1308
Db 1205 ESSPEKV---REEAASKGHVIEGKSGHILSYDNKNA-----REGTRSPRTAHMSL-K 1255
QY 1309 RTYDMMEGRVGRAS-----SASIEGLMGRATIPPERHSPH-HLKEQHHRIGSITQIGPR 1361
Db 1256 RSYEAVEGSIKQGMRESFVSAPLEGLICRALP--RGSPHSLKERTVLSGSIMQGTTPR 1313
QY 1362 SYVEAQEDYLREAKLLKREGTPPPPPPPRDLTEAYKTQALGPLKPAHEGLVATVKEA 1421
Db 1314 ATAESPEDGL-KYPKQIKRES-----PPIRAFEGAI-----TKGP-YDG-ITTIKEM 1358

Db 565 PHH---PQO---QHQQQQQQQQHQAHOQH---ORSSPAQQVQQQQQQNNSSRQSDNMC 616
Qy 68 -----LSEFPQNGERSQELHLPESHYSYLPGLCKSEMFIESKRPRELELLPDP 116
Db 617 RQVVTTPMGMLKAVETLPQQQKQHQHQHQHQ---QQQGRSQ-----PVSWSMTV 666
Qy 117 LLRP-----SPLLATQOPAGSEDLTKDORSULTKGLPVPSPPHPTDP 158
Db 667 VSQPVGTVTVTAGLSASHSGSGNVAAGLTGNTGSASTEAYHPQVEAISPTLP--SDS 724
Qy 159 ELELVPPRLSKBELIONMDVRDREITWVEQQISKLAKKQOQLBEA--AKPPEPEKPV- 215
Db 725 SIEERGRTSAKEDLLMQIKVDONEIKSAETMTTLRKKESLMEEAALAKEQRAAKELND 784
Qy 216 -----PPPIESKHS--LVQIYIDENRKAEEAHRILEGLGPQVE-----LPLYN 258
Db 785 NNNDQEPMVELSWRSQMLAEKIYAANKRTAQASHMLQNAADESSPGSVAGRPWMLPLYN 844
Qy 259 QPSDTRQYHENIKNOA--MRKKLILYFKRNHARKOWKQFCORYDQLMBALEKKYVERIE 317
Db 845 QPLDVEALAMLIRHOSQIRAPLLHILRKLKAERWAHNOQLVEKYTKDQADWQRCERME 904
Qy 318 NNPRAKESKYREYVEKOPPEIRKORELQERMQSRVGSGSLSAARSSEHVSEIID 377
Db 905 ASAKRKAREAKNREFPEKVFTELKQREDKERP--NRVGSR-----IKSEADLEEIMD 955
Qy 378 GLSEQNLSEKQROLAVIPMLYDADQOQRIKFTNMNGLMADPMKYDKQVNMVSEQEK 437
Db 956 GLOEQALEDKWRSYAVIPLMHADQRRCAYHNENGLIEDVAVHQORALNMTWTAGK 1015
Qy 438 ETRFKFMQHPNFGLIASFLERKTVAECVLYLTKQENYKSLVRSYRRGKSQOQO 497
Db 1016 ETEFKEYLQHPKNFGAIALSLDKSPQDCVRYLYLTKTENYKQLLRKSRQRTSRNPA 1075
Qy 498 QOQOQO-----QOQOQOQPMRPS---QEEKDEKEKE-----528
Db 1076 KAQAAPQCIIDSMTTGVNTRLQREOQKSGRSSAVERERAERAEVAEKAADA 1135
Qy 529 ---AEKEEKPE-----537
Db 1136 AKAAESAABKASAKATKAVEATAAGEKVAKAAAAAAAAATATTATTTSSSTSSSSSA 1195
Qy 538 -----VENDKEDLLK-----EKDDTDSG-----555
Db 1196 SSASTASSSTASPATLAGAADTKDAGKTASADKNAATAGGTATGTPTAATTATATA 1255
Qy 556 -----EDNDEKAVASKRKTANSQGRKRGRITRSMANBANSSEAITPQOSA-- 602
Db 1256 PPEISAGGEBAKNAEBAATAAGATVATAGTPATGASAAAGATATTATATAAAGK 1315
Qy 603 ---ELASMELNESSRWTEEE-----METAKKGLLEHGRNWSAIRMVGSKTVSQCKN 651
Db 1316 VKPETAPEPATGAAGADSRPDANDPLAKTASKAINAEGYN-----AIGNSSSSSSN 1369
Qy 652 -----FYFNYKKRQWLDE-----ILQOHK 670
Db 1370 ATGASAPVQGVTLNGFKPGVQTVVMANVKASCTGDDSDGANAGAAAGSLAATNATSG 1429
Qy 671 LKMEKERNARRKKKAPAAASEBAAPPPVVEDEMEASGV-----SGN-----713
Db 1430 DKIVKTTPTSSR---APNSTSTAA-----NESSGAGVNTYGHATTAGNYLGQKLKA 1479
Qy 714 ---EEMVYEEAEALHASQNEVPR-----GECSPAT--VNNSSDTEISPSHTEAAKDTG 763
Db 1480 AQVEGLGAGNELHSDVESKRRKRFELNSGEAGNATSAMTNSSTSGSMNISNHLKANA 1539
Qy 764 QNG-----PKPAT-----772
Db 1540 KQGSMMAKTSMATSSASVVTSTPSSASSSSLSASSMLLISAAVSMSTAAGNATSSSTAT 1599
Qy 773 -----LGADGPPPG-----PPTPRRTSRA--792

Db 1600 TTATASAIISLPLLAGDSGNSMVNANEILLADGDKKLASCFCVKCAEACPRTRPLPKGRGQ 1659
Qy 793 ---PIETPAS-----EATGAPTPPPAPPSPAPPV-----VPKKEKEETAAA 834
Db 1660 YGIPDETI PAGARVCNCSCKSVSRYPNCPPLPTCPNPKDRAQRLRNIPSLRFE---LA 1715
Qy 835 PVVEE---CEEQKPPAA-----EELAVDTGKA-----EEPVKSE 865
Db 1716 PEVRDPLMAEFOIIPPHATRCACSACLMRIRKLDLPQLNLTDGSSGAGSGGDETVDSTS 1775
Qy 866 CTEEEBEGPAKGDAAEATAEAGLKAEK-----895
Db 1776 SCDEREFG---GSDTASVESPEN--LQRHKSILTWKQKQOQOQOQOQOQOQOQOQ 1830
Qy 896 -----KEGSGRATTA-----906
Db 1831 SQPQPPPPAPQOQKSGSGRGDQGTPLIITPTRMSSKSGSGGAQTAGDNERLLPPAAGQA 1890
Qy 907 -KSSGAPQSDSATSACSADEVDE-----AEGDKNRLLSRP--SL 944
Db 1891 PKQKRTSEYDSDSATETADEENENSPANRQSPKVLPHGHGHGHGHANNVAGLQPPVANM 1950
Qy 945 LPTGDPPRANASPOK---PLDLKQLKQRAAAIPIQO--VTKVHEPPREDAAPTCKPAPP 999
Db 1951 GTGGGVQPCGAQOQVNGFISM-----RREAVNVQDCVFSVIERSLKHKGP-----1997
Qy 1000 PPNQNLQPSDAPQOPG-----SSPRGKSRSPAPPADKFAFAAEAKLPDPPCWT 1050
Db 1998 -----QPKGGQO 2043
Qy 1051 SGLPPVPVPREVIXASPHAPDPSAFSY-----APP---GHPLPLGLHDTARPVLPR 1098
Db 2044 -----ERKELTIVREYRQDPGILKQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2094
Qy 1099 PPTISNPPPLISSAKHPSVLERQIGAI SOGMSVQLHPVHYSEHAKAPVGPVTWGLPLPMDP 1158
Db 2095 PAAVAPPPP-----AHP-LTPTSTGCAGNNGTSDSLATLSVVNHSWMGVGIGHGPM 2148
Qy 1159 KKLAPFSGVKQQLSP--RQOAGPPESLGVPTAQAEASVL-RGTALGSVPGSGSTIKGIPST 1215
Db 2149 ASSAGIGYVDKATITPVKSSSGSSKSGSSASSHSTATPETI IYVVPVAHPQRGIPP- 2207
Qy 1216 RVPSSAI--TVRGSITHGTADVLKYGTITRI-IGEDSPSLDRDREGESLPGHVIYEG 1272
Db 2208 --PSQHSVHPAHPSHQHPAHPHSHQHTOLOVPEPEPOTLDSIKKPPRDCHSPHTG 2265
Qy 1273 KKHVLSYEGSMVTCCKEDGRSSS---GPHETAAPK---RTYDMMEGRVRAISSA 1325
Db 2266 -----AGSSSSSGSGSGSPSSDRHHGPPPTMSMKHIVRSGMYRGDTVTPVSLA 2316
Qy 1326 -----SIEGLMGRAIPPERHSPHHLKEQHHRGSIQTGIPRSYVEAQ-----ED 1369
Db 2317 APSSYLYPTRSVKTI GGGVVP-----GVLPGVPGSALYLQVPVPVPI 2360
Qy 1370 YLRREAKLLKRGTPPP--PPPSRDLTEAYKTOALGLPLKLP--AHEGLVATVKEAGRSI 1425
Db 2361 SISGQQLPPRAGQPPPAQPPSGRGVAKV-----PPKLSPOQAHLHPHSHGSHSPSOQ 2413
Qy 1426 HPIPREELRHTPELPLAPRPL--KEGSITQGTPLKTDGTGASTTGSKKHDSVRLSGSPRT 1483
Db 2414 QO 2469
Qy 1484 FPPVPHLDVMADARALERACEESLKS-RPGTASSSG--GSIARGAPVIVPELKGPR--Q 1538
Db 2470 LSKFPD-----GLVRQTTPEGVSGVPGGASGSGKSGSITQGTPLMP-----PHHLE 2517
Qy 1539 SPLTYEDHGAFAGHLPRGSPVTMREPTPLRQBSLSLSSKASQDRKLTSTPREIAKSPHS 1598
Db 2518 SKRPYESYKSKRSHSPAQQPGNQQLPPPQOQSS-----2552
Qy 1599 TYPHHHPHPISEYHLLRGVSDLYRSHIPLAFDFTSIPRGIPLDAAAAYLPRHAPN 1658
Db 2553 --FOAPP-----PQGYGVGVSSPY-----ARS 2572

QY 138 DRSITGKLEVPSPSPHTDPELELVPRLSKELIQNMDRVREITMVEQOISKLKKQ 197
DB 1 DRSITGKLEVPSPSPHTDPELELVPRLSKELIQNMDRVREITMVEQOISKLKKQ 60
QY 198 QOLEEAAKPEPEKVPSPPIESKHSLSVQIYIDENRKAABAAHRIEGLGQVQLPLY 257
DB 61 QOLEEAAKPEPEKVPSPPIESKHSLSVQIYIDENRKAABAAHRIEGLGQVQLPLY 120
QY 258 NQPSDTRQYHENIKINQAMRKKILYFKRNHARKQWKQFCQRYDQLEALEKKVERIE 317
DB 121 NQPSDTRQYHENIKINQAMRKKILYFKRNHARKQWKQFCQRYDQLEALEKKVERIE 180
QY 318 NNPR 321
DB 181 NNPR 184
RESULT 11
ABU70979
ID ABU70979 standard; protein; 224 AA.
XX
AC ABU70979;
XX
DT 10-JUN-2003 (first entry)
XX Human adipocyte Selected Interacting domain, SID, #610.
DE Human; prey; adipocyte; SID; selected interacting domain; anorectic;
KW antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS Homo sapiens.
XX
PN WO200286122-A2.
XX
PD 31-OCT-2002.
XX
PF 14-MAR-2002; 2002WO-EP003768.
XX
PR 14-MAR-2001; 2001US-0275734P.
XX (HYBR-) HYBRIGENICS.
XX
PI Legrain P, Daviet L;
XX
DR WPI; 2003-103412/09.
DR N-PSDB; ACAS7523.
XX
PT New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes.
XX
PS Claim 6; Page 308-309; 382pp; English.
XX
CC The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and a
CC record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are useful for
CC preventing or treating metabolic disorders such as obesity or diabetes.

CC The polynucleotides are useful as probes or primers. The complex is
CC particularly useful for identifying selected interacting domains (SID
CC (RTM)) for screening drugs that modulate the protein interaction, thus
CC exhibiting the therapeutic effect. The present sequence represents a SID
CC (prey) protein of the invention
XX
SQ Sequence 224 AA;
Query Match 5.9%; Score 774.5; DB 6; Length 224;
Best Local Similarity 73.6%; Pred. No. 1.3e-35;
Matches 159; Conservative 14; Mismatches 38; Indels 5; Gaps 3;
QY 116 PLLRSPPLATGQAGSEDLTKDRSLTGKLE-PVSPSPSPHTDPELELVPRLSKELIQ 174
DB 13 PLVHPLP---EGLRA-SADAKKDPAGGKHEAPSSISQPCDDQNASPSKLSKEELIQ 68
QY 175 NMDRVREITMVEQOISKLKKQOOLEEAAKPEPEKVPSPPIESKHSLSVQIYIDEN 234
DB 69 SMDRVREITAKVEQQILKLLKQOOLEEAAKPEPEKVPSPPIESKHSLSVQIYIDEN 128
QY 235 RKKAEEAAHRIEGLGQVQLPLYNOPSDTROYHENIKINQAMRKKILYFKRNHARKQ 294
DB 129 RKKAEEAAHRIEGLGQVQLPLYNOPSDTROYHENIKINQAMRKKILYFKRNHARKQ 188
QY 295 KQFCQRYDQLEALEKKVERIENNRRAKESKVR 330
DB 189 EQKICQRYDQLEALEKKVERIENNRRAKESKTR 224
RESULT 12
AAR99738
ID AAR99738 standard; protein; 619 AA.
XX
AC AAR99738;
XX
DT 27-SEP-1996 (first entry)
XX
DE Retinoid X receptor interacting protein RIP13.
XX Retinoid X receptor interacting protein; RXR; RIP; RIP13.
XX Mus sp.
XX WO9621677-A1.
XX
PD 18-JUL-1996.
XX
PF 08-DEC-1995; 95WO-US016311.
XX
PR 13-JAN-1995; 95US-00372652.
XX (GEHO) GEN HOSPITAL CORP.
XX
PI Moore D, Seol W, Choi H;
XX
DR WPI; 1996-342241/34.
DR N-PSDB; AAT31931.
XX
PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to modulate
PT or mediate RXR function, anti-RIP antibodies can be used to determine RIP
PT subcellular distribution patterns.
XX
PS Claim 2; Page 51-52; 90pp; English.
XX
CC Mouse retinoid X receptor (RXR) interacting protein RIP13 (AAR99738) is a
CC candidate transcriptional co-activator. It was identified using an in
CC vivo interaction trap system for the isolation of proteins that
CC physically interact with RXRs, esp. with the ligand binding domain of
CC human RXR alpha. Recombinant RIP13 can be obtd. using a cDNA clone
CC (AAT31931) obtd. from a mouse liver library. RIFs (see also AAR99735-37
CC and AAR99739) can be used to modulate or mediate RXR function, and may be
CC used therapeutically or to raise antibodies
XX

SQ	Sequence 619 AA;	
XX	Query Match	
XX	Best Local Similarity 5.4%; Score 718; DB 2; Length 619;	
XX	Matches 213; Conservative 76; Mismatches 212; Indels 90; Gaps 26;	
QY	1975 PSKSGSRPLV-----PPVGHATIAITPAKN--LAPHASDPAPPASADP 2021	XX
Db	70 PSQAQPHASVYVSEAGKDGPPKRYEELRTRGKTTITAAAFIDVITRIQIASDKA 129	XX
QY	2022 HREKTSQKFSQELRLSLGVHSGSYSPGVSPVSPSLTHDKGLPKHLELDKSH 2081	XX
Db	130 RERGSSDSS-----SSLSHRYETASDAIEVISPASSPAPQKQAYOPDMVKANQ 183	XX
QY	2082 LEGELRPKQPGPVKLGGEAHLPLR-----PLPESQSSSPLLQF--APGVKHORVVT 2134	XX
Db	184 AENESTROYEGP-----LHYSQOESPSQOQPLPPSSQSEGQVPRTHRLIT 234	XX
QY	2135 LAQHISEVITQDYTRHHPQOLGAPLPAPLYSPFGA--SCPVLDRRPPSDLYLPPDP--- 2189	XX
Db	235 LADHICQILITQFARN--QVSPQSTSTQTSALSSTPV---RTKTSRYSQSQST 289	XX
QY	2190 --HGAPA-RGSPH-----SEGGKRSPEPNKTSVLGGEDGIRPVSPGWTBPG-HSRSA 2240	XX
Db	290 VLHPRGPRVSPENLVDKSRGSRPGKSPERSHI---PSEPYEPISPQG---PAVHEKOD 343	XX
QY	2241 VYPLLYRDGEOTEPSRMGSKSGNTSQPPAFTSKLTESNMYKSKQKQIKNKLNTHNEN 2300	XX
Db	344 SMLLSQRGVDPAEQSDRSRPSGSIYLPSPFFTKL--ESTSPMKSKKQEIFRKINSGGG 402	XX
QY	2301 EPEYNTSQGTETFFNPAITGTGLMYRQAOVHASTNMGLEAITRKALMGKYDQWEE- 2359	XX
Db	403 DSDMAAQCTEIFNLPVAVTTGAVSSRSHSPADPAS--NLGLEDIIRKALMGSDDKVED 461	XX
QY	2360 -----SPPLSANAFNINLASALPAMPITAADGRSDHILTSPGGG-GKAKVSGRPSRK 2413	XX
Db	462 HGVMVHPV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKNSRK 513	XX
QY	2414 AKSPAPGLA--SGDRPPSVSVHSEGDGCRRTPLTNRVWEDRPSAGSTPPFVNPILMRL 2471	XX
Db	514 SKSPIQSGVLGTERSSSVSHSEGDYHQTTP--GWAMEDRPSSTGTFQFPYNPLTIRM 571	XX
QY	2472 QAGVMASPPPPGLPAGSGPL--AGPH---HANDEEPKPLCSQYETLSDSE 2517	XX
Db	572 ----LSSTPTQTACAPSAITQAAPHQONRIWEREPAPLLSAQYETLSDSD 618	XX
RESULT 13		
ABO07211		
ID	ABO07211 standard; protein; 3371 AA.	
XX	AC ABO07211;	
XX	XX	
DT	13-AUG-2003 (first entry)	
XX	Human p53 modifying protein, SEQ ID 171.	
DE	Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;	
KW	antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;	
KW	lung cancer; ovarian cancer; angiogenesis; cell cycle;	
KW	apoptotic disorder; cell proliferation disorder.	
XX	OS Homo sapiens.	
XX	PN WO200299122-A1.	
XX	12-DEC-2002.	
XX	03-JUN-2002; 2002WO-US017382.	
PF	05-JUN-2001; 2001US-0296076P.	
PR	10-OCT-2001; 2001US-0328605P.	
PR	15-FEB-2002; 2002US-0357253P.	

(EXEL-) EXELIXIS INC.

Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

WPI; 2003-156859/15.

N-PSDB; ACD13385.

Identifying modulators of the p53 pathway for use in treating apoptotic or cell proliferation disorders, comprises screening for agents that modulate activity of a human ortholog of genes that modify the p53 pathway in *Drosophila*.

Example 2; Page 505-515; 678pp; English.

The invention relates to identifying (M1) a candidate p53 pathway modulating agent, by contacting an assay system comprising a purified HM polypeptide (human ortholog of genes that modify the p53 pathway in *Drosophila*) or nucleic acid with a test agent under conditions, where but for the presence of the test agent, the system provides a reference activity, and detecting a test agent-biased activity of the assay system. Also included are modulating (M2) a p53 pathway of a cell (comprising contacting a cell defective in p53 function with a candidate modulator that specifically binds to a HM polypeptide comprising an HM amino acid sequence, where p53 function is restored), modulating (M3) a p53 pathway in a mammalian cell (comprising contacting the cell with an agent that specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4) a disease in a patient (comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with a probe for HM expression; (c) comparing the results with a control; and (d) determining whether the comparison indicates a likelihood disease). (M1) is useful for identifying modulators of the p53 pathway. A probe for HM expression is useful for diagnosing breast, colon, kidney, lung and ovarian cancer, in a patient, where the cancer has greater than 25 % expression level. Modulators identified by (M1) are useful in a variety of diagnostic and therapeutic applications, where disease or disorder prognosis is related to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell proliferation disorders (e.g. cancer). Another two new methods (M2 and M3) are useful for modulating the p53 pathway of a cell, thus restoring the p53 function of the cell, so that the cell undergoes normal proliferation or progression through the cell cycle. (M2) and (M3) are also useful for treating defects in the p53 pathway such as angiogenic, apoptotic or cell proliferation disorders. The present sequence represents a human p53 pathway modifying protein

Sequence 3371 AA;

Query Match 4.3%; Score 569.5; DB 6; Length 3371;

Best Local Similarity 19.6%; Pred. No. 1.6e-22;

Matches 568; Conservative 349; Mismatches 973; Indels 1011; Gaps 137;

QY 71 EFQPGNRSQELHLRPE-----SHSLPELGKS-----EMETESKR---PRLELPL 114

Db 849 ERKSGQKSHSVNTEKIGIDIDHTQSYRKQEQRRKQKQMEETAKSKFSPKDV-- 906

QY 115 DPLLRLPPLATQCPAGSEDLTKDRSLTKGLRFVSPSPPTDPELELYPRLSKBELIQ 174

Db 907 DEVERSLVHEVGKP--PDVTD-----SPPSKK-----K 935

QY 175 NMDRVDRDI-TWVQOISLKKKQOOLEEAAKPP-----EPEKVPSPPIES----- 221

Db 936 RMDHVDICTKRENRYSRQISEDSERTGSGSPSVRHGSHFDEDEPICSPLLSVKGSP 995

QY 222 -----KHSRLVQIYDENKK--AEAAHRILEGLPQOVELPLYNQPSDTRQ 265

Db 996 KYDEKVLPSYNTVREESLKFNPYSSRREQADWAKIKLSVNSEDELNRW-----DSQM 1051

QY 266 YHENTKINQAMRKLILYFKRRNHAKQWKQFCORYDQLMFALEKKVRIENNPR-- 322

Db 1052 KQDAGRFDVSFFNSII---KDSLRLKRSVRDL--EPGEVPSDSDGDEGSHKSHSPASAL 1105

QY 323 -----RAKESKVREYEEKQPEIKRKELQERMSRVQGRSGL-----SNSAAR 367

QY 2196 GS-----PH-----SEGGKRSPEPNKTSVLGGGEGD 2221
Db 2882 SEVLVMQSEYRLHPYTPROVIRIMVHPHTAVSEQRAADGVVKKVPASKAP-----QQPG 2937
QY 2222 IEVPSPEGHTPECHRSAY-----YLLLYRDGEQTPESRMGSKSPGNTSQ-----PP 2269
Db 2938 KEAAKTPDAKAAPTTPAPVPVPLPAPAPAPHGE-----ARILVTSPNQLQGLPLTPP 2993
QY 2270 APFSKLTESAMVKKQKQINKLWTHNRNE--PEYNI SQPCTEIFFNPAITGTGLMTY 2327
Db 2994 -----VVVTHGVQI-----VHSGELFQYRYGD-----IRTY 3021
QY 2328 RSQAQOE-----ASTNMGLEAIRKALMGKYDOWEE--SPPLSANAFNPLNASASLPAAMP 2382
Db 3022 HPPAQLTHTQFPAASVGLPSRTKTAAGPPPEGEPLQPPQVQSQPAPAPPCPPSQ- 3080
QY 2383 ITAADGRSDHLLTSPGGGKAK---VSGRPSSRKAKSPAPGLASG---DRPP-----SVSSV 2433
Db 3081 -----LQPGCGPFSKMPQVQSQAAGTGTGVEQFRLPAGPANRPPPEPHTQVORA 3129
QY 2434 HSB-GDCNRRTPLTNRVWEDRPSAGSTPPFPYVPLIMRLQAGVMASPPPPGLPAGSGPLA 2492
Db 3130 QAGTGTSPSPVSVSNKMDLPLVSLTQTAPKQPLFVPTTSG-----PSTPPG---LV 3179
QY 2493 GPHAWDEEPK-----ELLCSQ 2509
Db 3180 LPHTFQAPAKQDSSPHLTSQ 3200

RESULT 14

ABR47592
ID ABR47592 standard; protein; 3664 AA.
AC ABR47592;
XX
XX
DT 12-JUN-2003 (first entry)
XX
DE Breast cancer associated protein sequence SEQ ID NO:423.
XX
KW Human; breast cancer; cytostatic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2003004989-A2.
XX
PD 16-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019669.
XX
PR 21-JUN-2001; 2001US-0299887P.
PR 27-JUN-2001; 2001US-0301572P.
PR 18-JUL-2001; 2001US-0306501P.
PR 25-SEP-2001; 2001US-0325002P.
PR 05-MAR-2002; 2002US-0362585P.
PR 14-MAY-2002; 2002US-0380391P.
XX
XX (WILL-) MILLENIUM PHARM INC.

XX
XX Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kanatkar S;
PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
PI Bast RC, Hortobagyi GN, Puzstai L, Mexic F, Sahin A, Mills GB;
XX
XX WPI; 2003-210381/20.
DR N-PSDB; ACC50291.
XX
XX Breast cancer diagnosis or treatment by comparing the level of expression
PT of a marker in a patient sample with that in the control non-breast
PT cancer sample.
XX
XX Claim 1; SEQ ID NO 423; 128pp; English.

XX
XX The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level

CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences from
CC the present invention have cytostatic activities and can be used in gene
CC therapy. The method is useful for diagnosing and treating breast cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3664 AA;

Query Match 4.3%; Score 569.5; DB 6; Length 3664;
Best Local Similarity 19.6%; Pred. No. 1.8e-22;
Matches 568; Conservative 349; Mismatches 973; Indels 1011; Gaps 137;
QY 71 EFQGNERSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR---PRLELLP 114
Db 1142 ERKSQEKSHSVNTEKIGIDIDHTQSYRKQMEQSRKQKQEMETAKSEKFGSPKQDV-- 1199
QY 115 DPLLRPSPLLATGQAGSSEDLTKORSLTGKLEPVPSPPPPHPTDPELELVPPRLSKEELIQ 174
Db 1200 DEYERRSLVHEVGKP--PDVTTD-----SPPSKK-----K 1228
QY 175 NMDRVDRET-TMVEQOI SKLKKKQQLSEAAKPP-----EPEKVPSPPIES----- 221
Db 1229 RMDHVDFDICTKERNYRSSQISEDSERTGSGSVRHGSPHEDBDP GSPRLLSVKGSP 1288
QY 222 -----KHSRLVQI IYDENRKK--AEAARILEGLGPQVELPLYNQSDTRQ 265
Db 1289 KVDEKVLPSYNTIVRESLKFNPDYSSRREQWADWAKIKLSVLNSEDELNRW----DSQM 1344
QY 266 YHENIKINQAMRKLLILYKRRNHARKQWKFCORYDQLMEALKKVERIENRPR--- 322
Db 1345 KQDAGRFVDFVFNFI-----KRDLSLRKRSVRL--EPGEVPSDSDEGDEHKSHSPRASAL 1398
QY 323 -----RAKESKREYVEYKQFPEIRKQRELOERMQSRVQGRSGI-----SMSAAR 367
Db 1399 YESSRLSFLLRDREKLRERDERLSSSLERNKFPYFALDKTITPTTKALLERAKSLSSSR 1458
QY 368 SEHEVSEIIDLGE-----QENLEKQMRQLAVIPPMYLDADQQRIKFINMGLMADPMKYV 423
Db 1459 EEN--WSFLWDSDRFANFRNNKDEKVD SAPRIPSWYMKKKIR-TDSEGRMDKKEDH 1515
QY 424 KDRQVMNMWSEQEKETFREKFMQHPNFGLTASFLEKRTVAECVLYYLTKNENYKSLV 483
Db 1516 KEE-----QERQLFASRL-HSSI FEQDSKRLQ-----HLERKEEDSDFTS 1557
QY 484 RRSYRRRGKSQQQQQQQQQQQQQQQQ-----MPRSSQEKDEKEKEKEKEE 533
Db 1558 GRIY--GK---QTSEGANSTTDSIQEPVVLPHSRFMELTMRQOKEKEDQPKPEVKEQ 1611
QY 534 EKPEVNDKEDLLKEKTDGTDGENDKEAVASKGRKTANSQGRKGRITRSMANEANSE 593
Db 1612 ---DTE-----HPKTPESAPENK-----SELKTP 1634
QY 594 EAITQQQAELASLMELNESSRWTEEMETAKGLLEHGNWSAIARMVGSKTVSQCKNFPY 653
Db 1635 PSVGPPSV-----TVVLTESAP-----SALEKTTGDKTV----- 1663
QY 654 FNYKQRQNLDEILQQLKLMKERNARRKKKAPAAASEAAFPVPVDEEEMASGVSGN 713
Db 1664 -----EAPLVTEKTVEPATVSEAKPASEPAPA 1692
QY 714 EEMVVEEALHASGNEVPRGCSGPATVNNSSDTEIPSPHTEAKDQTGQNGPKPPATL 773
Db 1693 PVEQLEQV-----DLPPGA-----DPDKEAAMP--AGVEEGSGDQPP-YL 1731
QY 774 GADGPPGPPPPPPRPTSRAP--IEPTPASEATGATPPAPPPSPAPPPVVPVKEKEEET 831
Db 1732 DAK-----PPTGASFQSAENSVDPEDPS-----TQPLSKPAQKSEANEPKPKPDAT 1780

XX	Homo sapiens.
OS	
XX	WO300058473-A2.
PN	
XX	05-OCT-2000.
PD	
XX	
PF	31-MAR-2000; 200OWO-US008621.
XX	
PR	31-MAR-1999; 99US-0127607P.
PR	02-APR-1999; 99US-0127636P.
PR	05-APR-1999; 99US-0127728P.
PR	30-MAR-2000; 200OUS-00540763.
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
DR	N-PSDB; AAC76700.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease.
XX	
PS	Claim 11; Page 3700-3708; 5507pp; English.
XX	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulneryary;
CC	antiparitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC	anticongulant; antiarthritic; immunosuppressant; immunostimulant;
CC	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC	dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC	antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC	sequences can be used for determining the presence of or predisposition
CC	to, or preventing or treating pathological conditions associated with an
CC	ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC	proteins in gene therapy vectors. The proteins and nucleic acids may be
CC	used to treat cancers, proliferative disorders, neurodegenerative
CC	disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC	diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC	storage, systemic lupus erythematosus, severe combined immunodeficiency
CC	(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC	disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC	cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC	enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX	
SQ	Sequence 3266 AA;
	Query Match 4.3%; Score 565.5; DB 3; Length 3266;
	Best Local Similarity 19.5%, Pred.No.2.5e-22;
	Matches 567; Conservative 350; Mismatches 973; Indels 1011; Gaps 137
QY	71 EFQPCNERSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR---PRLLELP 114
Db	739 ERKSGQEKSHSVNTEEKIGDIDHTQSYRKQMEQRKKQOMEMETAKSEKTGSPKDV-- 796
QY	115 DPLLRPGPLLATGPAGSEDITKORSLSGTLKEPVSPSPPHPTDPLEILVPRLSKEELIQ 174
Db	797 DEYERRSLVHEVGKP-PQDVTDD-----SPPSKK-----K 825
QY	175 NMDRVDDRRI-TWVRQQISKLLKQQQLBEAAKPP-----EPEKPVSPPIPES----- 221
Db	826 RMDHVDFDICTRRNRYNSRSQISDSERTGSPSVRHGSFHEDBDIPGSPLLSVKGPS 885
QY	222 -----KHRSVLQIIVDENRKK--AEAHRILEGLGPQVELPLYNQSDTRQ 265
Db	886 KVDEKVLPSYNITVREESLKFNFPYSSRREQWADNAKIKLVINSEDELNRW----DSQM 941
QY	266 YHENIKINQAMRKILIFYKRNHARKQWKFCORYQDLMEALBKVKVERIENRP --- 322
Db	942 QKDAGRFDVSPFNGII-----KRDLSRKESVRLD--EPGEVPSDSDEGDGHKSHSPRASAL 995

CC V(D)J Res) activity or that modulates the interaction between KRC and a
CC TRAF molecule, so that inflammation or apoptosis in the cell is
CC modulated. Also provided are methods for modulating immune cell
CC proliferation or activation and for inhibiting metastatic growth of a
CC tumour cell. The method is useful in modulating an immune response, such
CC as inflammation or apoptosis, using agents that modulate KRC activity.
CC The method may also be used in identifying agents that modulate KRC
CC activity which can be used for treating or preventing disorders
CC associated with an aberrant cell proliferation or survival, such as
CC autoimmune disorders or graft-versus-host disease. The present sequence
CC represents a human KRC protein, a DNA binding protein. KRC is a member of
CC zinc finger proteins that bind to the kappaB motif
XX
SQ Sequence 2406 AA;

Query Match 4.2%; Score 552.5; DB 6; Length 2406;
Best Local Similarity 19.8%; Pred. No. 9.2e-22;
Matches 553; Conservative 323; Mismatches 966; Indels 953; Gaps 136;
3 GSTQLVAQTWRATEPRYPHSLSYVQIARTHTDVGLEYQHHSRDYASHLSPGS---II 59
69 GSQKGTGQ--QCKPKRPPIEASVHI---SHVPOHPLTPAFMSPCKPHEHLEGSTWQLV 122
60 QQRRRP--SLSEFQPG--NERSQELHLPESHVLP-----ELGKS 98
123 SPWRLGPGSGLLA---PGLHPQSQ---LLPFSHASIIPEDLPQVFKVFPVPRPSQVSLKPT 176
99 EMEFTIESKPR-----LELLPDPLLRPSLLA---TQO-----PAGESDLTKD--- 138
177 EBAHKKERPKQPKGYICQYCRPCAKPSVLQKHRSHTGERPYPCGCGFCKTKSNLY 236
139 -----RSLTGKLEPVSPPPHPTDPELELVPPRLSKEELIQNMDRVDREITWVEQOI 190
237 KHRKSHARIKAGLASGMGEMYPH--GLEMERIPGEEFEP-----TEGESTDSEET 288
191 SKLKKKQOOLEBAAKPPPEKVPSPPIES--KHRSVLQIYDENRKAHAHRIEGL 248
289 SATSGHPAELS-----PRKQPLSSGLYSGSHSSHERCSLSQSTAQSLDPPPPFV 342
249 GPQVELPLNOPSDDTQYHENIKINAMRKLI-----LYFKRRNHAK 292
343 EFSSEHPLSHKPDHTTIQKALRLSERKKVDEQAFSPGSKGSTESGYFSRESAEQ 402
293 QMKQKFCQRYDQLEALEKVERIENNNRRRAKESKVREYIEKQFPPIRKQRELOERMOS 352
403 Q-----VSPPTNAK-----SYAEIIFGKC-----G 423
353 RVGQSGSLMSGAAR-----SEHEVS-----EIIDGLSEQENLEKQMRQLAVIPP 397
424 RIGQRTAMLTATSTQPLPLSTEDKPSLVPSTQVIEHTIKLTINE-----AVVDT 478
398 MLYDADQQRKIFNNMGLMADPMVYKDRQVMMNSEQEKETFREKFMQHPKNGFLIASF 457
479 SEIDSVKPRRSLRRSSMESP--KSLYREPLSSHSEKTKPQOSLLSLQHPSTAPPVPL 537
458 LERKTV--AECVL-----YYILTKKENYKSLVRSY-----RRRKSOQOQOQOQO 502
538 LRSMSPSAACTISTPHFPFRGSYSPDDHITDSEALSRSSHVFTSHPRMLKQPAIEPL 597
503 QOQOQOQOQPMRPSQOEKE 562
598 GGEYSSEEPGSS---KDTASKPSD-----EVE--PKSEELATKKT----- 632
563 AVASKRKTANSQGRKG-----RITRSMANSEEAITPQOSAFSLASMELNESRWTE 617
633 ----KGLKLT-----KGVIECNIGARYKRDNYEAHKKYCYSELQIAKPIAGTHTS 682
618 BEMETAKGLLHGRNWSAIAIRMGVSKTVSCKNFYFNKRNLDLILQOHLKMEKER 677
683 PE---AEKQIEH--EPWSQ-----MMHYKLGJGTTLEL 709
678 NARRKKKAPAAASEAAFPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGCS 737

710 TPLRKRKRKESLGDEE--PPAFESTKQFQS-----P.740
738 GPATVNNSSDTEISIPSPHTEAAKDTGQNGPKPPATLGADGPPGPPPTP-----PRRT 789
741 GFSDAARNLPLESTKSP--AEPSKSV-----PSLEGTGFOPRTPKPGSGSESGKERT 792
790 SR---APIEPTPASBATGA-----PTPPAP-----PS 814
793 TSKEISVIQHTSSFEKSDSLEQPSGLEGDKPLAQFPSPPPAPHGRSAHSLQPKLVQRN 852
815 PSAPPVVPKKEKETAAAPVVEGEQKPPAAABELAVDTCKABEPVKSECTEE---ABE 872
853 IQVPEILVTEEDREDTPEPEPPKEPEK-----TEEFQWOPORSQTLAQLPABK 900
873 GFAKGD---AEAAEATAEGALKAEKGGSGRATTAKSSGAPOQSDSATCSADEVDEA 929
901 APPKKRRLRLAEMAQSGSESSPSSVP---LSRSPQESNVSLSGSSRSASFERDDHGKA 957
930 EGGDKNRLSPRP---SLLTPTGDP---RANASPOKP-----LDLKLKORAAAI 975
958 EAPDPSSDMRPKPLGTHMLTVPSHHPHAREMRSSASEQSPNVSHSAHMTETRSKFDYGS 1017
976 IQVTKVHPPREDAAATPKAPPAP---PPQVLO-----PESD---AP--QO 1014
1018 LSLT-----GFSAPAVPAPAGEAPPRERKCFLVRSPSLRPPSESELEVPKGRQ 1067
1015 PGSSPRGSRSPAPPADKEAFAAEAQKLPDGPCTWGLPPP---VPPREVKA--- 1065
1068 ESEEPQSSSKPSAKSSLSQISSAATSHGGPGGPGQDRPALGPTVYVYTAALQVHHIP 1127
1066 -----SPHAPDP--SAFSA---PFGH-----PLPLGLHD 1090
1128 VAQTLHEKPYLPVPVSLFSFOHLVQHEPGSPPEFFSQAMSSLLSSPYMPPPLPSLFQ 1187
1091 TARPVLPRPTISNP-----PLISAKHPSVLEROI---GALSQK----- 1129
1188 A--PPLPLQPTVLHPGQHLPLQMPHPANI--PPRQPPSLPMPYPTSSALSGFFFLPQSO 1245
1130 -----SVQLHVPYSEHAKAPVGTWGLPLPMDPKKLAPPSPGVKQE--QLSPRGQAGPP 1181
1246 FALQPLPGDVESHLPQIKTSLAPLATGSAG-----LSPSQEYSSDIRLPP---VAP 1293
1182 ESLGVPYTAQAEASVLRGTALGVPGGSITKGIP---STRVPSDAITYRG-----SITHG 1232
1294 ASSAPTAP-----LALPACPDWVSLVVRVQTNMPSYGSAMYTTLISQILVTQSO 1348
1233 TPADVLY-----KGTITRIIGED-----SPSELORGREDSLPKGH---VIVEGK 1274
1349 SSATVALPKFEPSPSKG---TTVCADVHEVGPSPGLSEEQSRAPPTPYLRVPVTLPERK 1406
1275 GHVLSEVGMSVTQCSKEDGRSSSGPPHETAPKRTYDM-----MEGRVGRASIASIEG 1329
1407 GTSLSSESITSL-----EGSSSTAGSKXVLSPAGSLELTMTQOQKRVKEEASKADEK 1461
1330 LMGRAIIP-----ERHSPHLKQHHIRGISTQIPSPRSVEAQEDYLRREAKLL 1378
1462 L--ELVKPCSVLVTSTEDGKREPKSHL-----GNQOG-----RRELEML 1499
1379 K-----RECTPPPPPSRDLTEAYKTAQALGPLKL---KPAHEGLVATVKEAGRSIHE 1427
1500 SSLSDPSDSTKEIPPLPHALSHGOAPGEALKEYPQSGKPHRRGLTLP-----S 1550
1428 IPRESLRHTPELP--LAP---RPLKEGSIQTGTPLKYDTGASTTGSKKHVDVRLISGPGRT 1483
1551 VKEDSKQPDPLSLAPSSLPSE---TSSRPKASQEG---TDSKK-----VLQ 1594
1484 FPPVH-----PLDVM--ADARALERA---CYEESLKSRPGTASSGGSIARGA 1526
1595 FPSLHTTNNVSMCYLNIKPNHIQHADRSSVYAGWCISLNPPLPGVSTKAALSLLR--- 1652
1527 PVIVELGKPROSLTYEDHGAPFAGHLPGRSPVTWREPTPELOE---GSLSSSKASQDR 1583
1653 -----SKQVKSKETYTWATAP---HPEAGRLVSSSRKPRNTEVHLPLSLVSPGQKDL 1702

QY 1584 KLTSTPREIAKSHSTVP---EHPHPISPYEHLRGVGVLDYRSHIPLAF-----1632
Db ARVEKEERERGEPEADAPASQORGEPAKIFE-----GG---YKSNEEYVYVRGRGRGK 1753
QY 1633 -----DPTSTIPRGIPLDAAAYVLPRL-----APNPTYPHYPPYVIRGYPT 1676
Db YVCEEGICRCKPMSLKHITHTDVRYVYKCHFAFTKGNLTGHMKSKAHKCKCQET 1813
QY 1677 AALENRQTIINDYITSQMHNTATAMAQRADMLRGLSPRESSALNAYAGPRGIIDLQ 1736
Db 1814 GVLE-----ELEAEBSGTS-----DDLFDQSEGESEAV-----1842
QY 1737 VPHLPVLVPTPTATAMDRLAYLPTAPQPFSSRHSSPLSPGPTHLTPTTTSSSR 1796
Db 1843 -----EEHQFSDL-----EDSDS 1855
QY 1797 ERDRERDRDREREKILTSTTT-----VEHAPIWRP-----GTQSSG 1836
Db 1856 DSDLDDEDEDEESQDELRSFSEAPPGPPHALRADSPILGPPDPAPASGTEATRG 1915
QY 1837 SSGS-----SGGGGSSSRPA-----SHSHAHQHSPISPR-----1866
Db 1916 SSVSEAERLTASSCSMSSQMFGLPWLGPAPLGSEVKGDTGSALSYPVSPRRPWSPSKEA 1975
QY 1867 -TODALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPV-----1907
Db 1976 GSRPPLARKHSLTKNDS-----SPQRCSPAREPOASAPSPGLHVPDGRGMGLPCG 2027
QY 1908 RPAATFPATHCPLG-----GTLGVYPTLMPEVLLPKEAP-----RVA 1946
Db 2028 SPRLQSPITLPLGLRELAPRAHVLSKLEGTDT---PG-----LPRYSPTRRWSPGQAB 2078
QY 1947 RPERPRADTGHAPLAKPAPARSGLPASPSSKSGE-----PRPLVPVVS-GHATARTP 1998
Db 2079 SPPR-----SAPPCKWALAGPSPSAGHGGLGAPVLPFPAPLPHKLSRSP 2128
QY 1999 AKNLAPHASPDPPAPASADPHREKTKOSKPFISOELRLSLGYHGSYSPEGVSPVSP 2058
Db 2129 ETCASPWOKA-ESRSFSPGPAH--PLSSRPFS-----ALHDPHGHLARTENI-----2176
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGELRPQKGP-VKLGGEAAHLPHLRPLPESOPSS 2117
Db 2177 -----FSH---LPLHSQHLTRA-----PCPLPIPG-----IQMVQARPGAH 2210
QY 2118 PLLQATAP-----GVGHRQVVTLAOHISEVITODVTRHHPOOLSAPLPAP---LYSFGP 2168
Db 2211 PTLPGPTAAWVGFGGGSDLTGARE-----AQERGRWSPTESSASVSPVAKVSKFTL 2265
QY 2169 ASCPVLDLRRPDLPLPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEGIEPVSP 2228
Db 2266 SS--ELEGDRDYPKERERTGGGPRPDWTPHGTAPEAETFTHS-----PCTPP 2312
QY 2229 EGMTEPCHRSASVPLLYRDEGOTEPGRMSKSPGN---TSQPPAFESK-----LTSNSGA 2281
Db 2313 DTLPREPQGR-----RAAQWSRPLSPAPANPEFSATPPLDRSSVSGCLAEA-SA 2363
QY 2282 MVKSKQEIKNKLNTH-----NRNEPEVNIQOP 2309
Db 2364 RPPARTRNLSGESRTRQDSPKPSGSGEPRAHPHQ 2398

RESULT 18

ABB62819
ID ABB62819 standard; protein; 1963 AA.
XX ABB62819;
XX ABB62819;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 15249.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL06922.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX Disclosure; SEQ ID NO 15249; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1963 AA;
SQ Query Match 4.2%; Score 550; DB 4; Length 1963;
Best Local Similarity 19.9%; Pred. No. 9.7e-22;
Matches 486; Conservative 233; Mismatches 824; Indels 898; Gaps 110;
QY 343 QRELQRMQSRVQGRSG---LSMSAARSEHEVSEIIDLSEQENL-----EKQMRQLAV 394
Db 20 ERELESRS--PGWADGDLMLFLRAARSMAAPQGMCDGLEDGCLAAASRDDTTINALDV 78
QY 395 IPEMLYDADQORIKFINMGLMADPMKYKDRQVMMWSEOEKETPREKPMQHPKFGLI 454
Db 79 LHDSGYDPGKA-----LQALVKCPVSKGIDKK---WTEDETKFKIKGLRQFGKFFRI 128
QY 455 -ASFLEKTKVAECVLYYLTKNENYKSLVRSYRRGKSQOQOQOQOQOQOQOQOQOQO 513
Db 129 HKDLLPHKDTPELVFYLYWKTPGANN--NRPHRRRSALARNRVRTRANNSNTP-- 184
QY 514 RSSQEEKDEKEKEAEKEEKEPEVENDEKDLKXETDTSGBDNDEKEAVASGKRKTAN 573
Db 185 -----PKKEDTPEPT-----ATTATAATAA 206
QY 574 SQGRKGRITRSMANEAISEAITPOQSAELASMELESSESRWTEEE-METAKKGLLEHGR 632
Db 207 SE-----TASRSSPAVSKE-----ENSLTEDDASECDSDSLTHKR 243
QY 633 NMSAIARMVGSKTVSQCKNFYFNKKRONLDELQOHLKMEKERNARRKKKAPAAASE 692
Db 244 DESP-SRM-----RTRN-----KQNNNSTSSGNNTAGN 272
QY 693 EAAFPVVEDEEMESGVSGNEEMVVEAEALHASGNEVPRG-----ECSGPATVNNSSDT 748
Db 273 GGNATSISSGSTGGGAAGSSSSKQDSANAV-ANGKRPKRGSSETPDVSGGASVDS--- 327
QY 749 ESIPSPHTAAKDTGONGPKPATLGADGPPGPPPTPPRPTSPAPTEPTPASEATCAPTP 808

Db 328 ---PKPTTKAVABESSANKK-----GGKQDTPNKKRTEQSENEPS-AHEENAIKEK 375
QY 809 PPAPSP-----SAPPVVPK-BEKEBETAAAPVEGEQKPPAAAEALAVDTGKAE 860
Db 376 RRPDSFVESMNSDRPDSVLDGSENTDTTTA-----EQSTKDSKE-TVSCKEERE 428
QY 861 PVKSECTEAEAGPAKGADEA-AEATAEALKAEEKGSGRATTAKSSGAPOQSDSS- 918
Db 429 MVTNDLEAKAE--KAIRAEALAEBSKDSAIKNMDEE-----TNIQAPSSADTSL 476
QY 919 -----ATCSA--DEVDEABGDKNRLSPRSLLTPTGDR 952
Db 477 VDPGNALPSPVAAPITMKVETIATVEALNASVDRKEAIERMESCDSPMLKLATIK 536
QY 953 ANASPOKPLDLKOLKORAAAIPIQTVKHEPPREDAATK-PAPPAPPPQN----- 1004
Db 537 QEVSPQOQHMQOQSO-----QMQOOLAPVGIPOPPSCPPSESVIKKEP 582
QY 1005 LOPESDAPQPGSSPRGKSRPAPPADKEAFAAEAQKLPDPPCWTSGLPFPVPPREVIK 1064
Db 583 MEDSDMATCNQNSNEPQDLKVIEKNEALKHSAGGLPPSGPC----- 626
QY 1065 ASHPADPSAFVAPPG--HPIPLGLHDTARPVLPRPTISNPPPLISSAKHPSVLEROI 1122
Db 627 -----APPSALHPL-----SGAP--VESQBFPLHQ----- 650
QY 1123 GAISQMSVQLHVPISE-HAKAPVGPVTMGLPLPMDPKKLAPFSGVKQQLSPRQAGPP 1181
Db 651 -----HMPHGQVTTQPPGVLIIDG-----QKYGPSGGQVPP 682
QY 1182 ESLGVPTAQEASVLRGTALGSPVGGISITKIGIST--RVPSDSAITVRGSIHTGTADVLV 1239
Db 683 -----QPQLHSDAAGVSGA--PPGAPTTPQKYPPEMEMKE-----AQDLKY 724
QY 1240 KGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHLVSYEGGMSVTOCSKEDGRSSG 1299
Db 725 -----PPPPPLD-----ALKYQEMQAAAA----- 745
QY 1300 PHETAAPKRTYDM-----MEGRVGRAISSASTIEGLMGRAIPEP----- 1339
Db 746 -----AAAAGKDKMYMEEQQKYNVELSAH-----QPPSPGYQDSLKIDIKPG 792
QY 1340 ---HSPHLKEQHIRGSIITQGISPRVSEAQEDYLREAKLLKREGTPPPPPPSRLDTEAY 1397
Db 793 FGHLPNV-----GSP--LDAAHKY-----GPPFTSQESQOQ 822
QY 1398 KYQALGPLKKAHEGLVATVEAGRSIHEIPEELRHPTPELPLAPRP-----LKEG- 1449
Db 823 QPQ-----PPAHQVPPGATPPPGIA--MPKPHYQHDVQTPPLGRPFPEPTGLMLKYGD 872
QY 1450 --SITQGT- -LKVD---TGASTTSGKXHDVRSILIGSPGRTFPPVHPLDVMDARALER 1501
Db 873 PLAAGVPPQDLKYMPMPVVSQAGPADVKYGENLIKSSPYGPPESPIDA----- 923
QY 1502 ACYBSLKRPGTASSGSGSIARGAPVIIVPELGPQRPSLTLYEDHGAFFAGHLPRG----- 1557
Db 924 ----SARSTPG--QDSQGSNSNSQPSMPPQPFQFSHPHSPMPSAGGGLPPGMHPQ 976
QY 1558 -----SPVTWREPT-----PRLOEGLSSSKASQDRKLSTPRE 1591
Db 977 NIHGPPPGAAGSGSQPPPPPTTSHQPTTSAGPSPLOHGLHPGHQHSQSVASSIPPS 1036
QY 1592 IAKSPH--STVPEHHHP-ISPVEHLLRGVGVDLVYRSHIPLAFDPTSIPRGIPLDMAAA 1648
Db 1037 STIGIPTLTMAPSHMHPHLPHAH-LQGL-----HRP----- 1068
QY 1649 YYLPRHLAPNTPHYLPYLLIRGVPTDAALENROTIINDYITTSQOMHNTATAMAQRAD 1708
Db 1069 ----HDLPPSMHPAPMPLSLQCHP-----QHGHGLPPSHTSQOQOQOQOQOQPGG 1118
QY 1709 MLRGISPRES-----SLAINVAAGRGIIIDLSQVPH-----LPVL 1743
Db 1119 TVRTSPAQQPPRSMHDPOSSREPPTSPQSTTTMAGSSGPG-PPPQSPHAHRTSPPLGL 1177

QY 1744 V---PPTPGT---PATAMDRLAYLPTA-----POPFSSRHSSSPSPSG----- 1781
Db 1178 AGSGPPPGGLIGHPMAIHPHLAHLPPGPAHAALAHPGHLLSHIAGLCPGGGPALLA 1237
QY 1782 -----PTHL-----TKPTTTSSRERDRDRDRDREREKSIILTS 1817
Db 1238 GFGGLGIPESALSRRTPPSHLPHSHASSAPUTAHSA-----SMTSTSMST 1285
QY 1818 TTTVHAIWR--PGTEOSSSGSSG-----GGG-----SSSRPASHSHAH----- 1858
Db 1286 TSTVFPASSPASRSPVOISSGGSPGSPGPMNPSSAAAAAHHRAASPASSVS 1345
QY 1859 ---QHSIPRPTQDALQORPSVLHNTGMKGIITAV-EPKPTVLSTSTSSVPRPAATPP 1914
Db 1346 SLRSRQSLHPVQSPSLSHPS---SSALSAAAAVAERDRHALMRQOS-----PHMTTP 1396
QY 1915 PATHCPL--GGTLGQVYP-----TLMPEVLLPKEAPVAR-PERRPADTGHAPLAKP 1963
Db 1397 PVSNASLMAPLSKMYAPQPGQGLGTGTPPHLRPGASPPVIRHPQMPL----- 1445
QY 1964 PARSGLEPASSKSGSERPLVPVSGHATIAITAKNLAHPHASDPDPAPASADPHR 2023
Db 1446 -----PLPIAPGGGIPQIGVHPGQSPYPH-----PULHPSVFSYPHH 1483
QY 2024 EKTQSKPFSIQBELRSLGYHGSSYSPGVEPVSVPSPSLTHDKGLPKHLEELDKSHLE 2083
Db 1484 H-----PFN-----SPGYAPYP-----GFAYMKP----- 1505
QY 2084 GELRPQKPGVKLG- EAAHLPHLRPLP-----ESQSSSPLLQTAGVKGHORVVTLAQ 2137
Db 1506 ----FPQGLDPAAVMAAHAGLQPPPPQMRQDEQNAAAAAQAAAEKQHAAAAA 1561
QY 2138 HISEVITQDY-----TRHHQOOLSAPLAPLYSFPFGASCPLDLRRPDSL----- 2183
Db 1562 QOHKAPQOQPGGMPNKPPTPKTPQPGGMP-----PGMGPGTPTGLPFGAYPGSH 1615
QY 2184 ---YLPPDPHG--APARGSPHS-----EGGKRSPEPNKTSVLGGGSDGTEPV----- 2225
Db 1616 MPGYQGPFGSPFPAPQDQGHGLKPTSHMDALRAHAHANSAGMGGGHHPTPLIDIE 1675
QY 2226 -----SP-----PEGM-----TEFGHSRAVYPLLYVRDGHQTEPSRMGSKSPGNT 2265
Db 1676 PDPEEIPSPTHNIPRGSPSPAKPDDTECHRSQSAIFVRHIDRGDYNSTR----- 1726
QY 2266 SQPPAFFSKLTESNAMYKSKQEKINKLNTNRNEPEVNIQSOPGTEIFNMFAITGTGLM 2325
Db 1727 --TDLIFKPVADSKLA---RKREERDRKLAERERROOQOQO----- 1764
QY 2326 TYRSQAVQEHASTNNGLEAIRKALMGKYDQWEEESPILSAN-----AFNPL 2371
Db 1765 --QOQOQOQAAAAQAAQAAKMA-----ELKPPVADTALQSLSEYARPHVAFSPV 1814
QY 2372 -----NASAL-----PAAMPITTAADR 2389
Db 1815 EQMVPYHHPMGPMYRERELEIKNAQAAAAQSRLDPHWMYRRGIHPSOPFLYANPAI 1874
QY 2390 SDHTLTSGGGGKAKVSGRPSRKAASAPGLASGDRPPSV 2430
Db 1875 SOMERERIGIPPHHVGLDPGEHVMVMPQPPAGFQLPPNV 1915

RESULT 19

ABU08487

ID ABU08487 standard; protein; 8991 AA.

XX ABU08487;

XX 24-JUN-2003 (first entry)

DE S. pneumoniae pneumococcal surface protein A (PspA) protein.

XX Pneumococcal surface protein C; PspC; pneumococcal surface protein A;

KW		alpha-helical; proline rich; repeat region; pneumococcal infection; PspA;
XX		antibacterial.
XX		Streptococcus pneumoniae.
OS		Key Location/Qualifiers
FH		Misc-difference 1.8991
FT		/note= "All Xaa residues within this sequence are
FT		unknown"
XX		US65006113-B1.
XX		31-DEC-2002.
PD		16-SEP-1996; 96US-00714741.
PF		15-SEP-1995; 95US-00529055.
XX		(UYAL-) UNIV ALABAMA.
PA		Briles DE, McDaniel LS, Swiatlo E, Yother J, Crain MJ;
PI		Hollingshead S, Tart R, Brooks-Walter A;
XX		WPI; 2003-361534/34.
DR		Isolated PspC amino acid sequence used as polymerase chain reaction or
PT		hybridization probe, comprises pneumococcal surface protein having alpha-
PT		helical, proline rich and repeat regions.
XX		Disclosure; Col 145-188; 186pp; English.
XX		The present invention relates to the isolation of Streptococcus
CC		pneumoniae pneumococcal surface protein C (PspC), and the polynucleotide
CC		sequence encoding it. PspC is a pneumococcal surface protein A (PspA)-
CC		like protein having alpha-helical, proline rich and repeat regions. The
CC		PspC and PspA proteins may be used in a vaccine to protect against
CC		pneumococcal infections. The polynucleotide sequences encoding PspC and
CC		PspA may be used for the expression of the proteins, and as PCR primers
CC		or hybridisation probes. The present sequence represents S. pneumoniae
CC		PspA protein
XX		Sequence 8991 AA;
SQ		Query Match 4.1%; Score 539.5; DB 6; Length 8991;
		Best Local Similarity 18.5%; Pred. No. 2.8e-20;
		Matches 553; Conservative 312; Mismatches 1122; Indels 1001; Gaps 116;
QY	116	PLLRPSLLATGQPAGSDEITKRSRLTGKLEVPSPSPHTDPELELVPPRLSK-EELIQ 174
Db	5314	PAPAPKPDELKEIDSDSEDYVKE---GFRAPL-----QSELDAAQAKLKLELSD 5361
QY	175	NMDRVDREITMVEQI-----SKLKKQQOLEE 202
Db	5362	KIDELDREIAKLELDQLKAEBENNVEDYFKEGLEKTKIAAKAELEKTEADLKAVNSPEK 5421
QY	203	EAAKPPEPKVPSPPTTESKHSRLVQIIYDENRKKAEEAHRILEGLGPVELPLYNOPS 262
Db	5422	PAEPSQEPKPAEAPEQP-----TEPTQEPKPAEQOPAPAPOPEKPAETTPAP 5473
QY	263	TROYHENIKINQAMRKKLILIFYKRNRHARKQWKQFCQR----YDLMEALEKKVERIEN 318
Db	5474	KPE-----KPAEQPKAEKPADQCAEDYARRSEEEYNRLTQQOPPCKAEKPAP 5520
QY	319	NPRRRAKESKVREY-----YEKQF--PEI-----RKORELOERMQRVRG 355
Db	5521	APTKGGSALDQEAAAPHQVADLEKQITGPEIFLGADPEADTAARPNELAAK-QELA 5579
QY	356	QRSGSL-----SMSAARSEHEVSIIIDL-SEQENLEKOMROLAVIPML 399
Db	5580	QKPTGLEKLLDSLDPGGKTQDELKEAGEAELDKADELPKNKVDLKEISNLEI---LL 5636
QY	400	YDAQOQRIKFINMGLMADPMKVYIKRQVMNMWSEQEKETFREFKFMQHKNFGLIASFLE 459

CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3080 AA;

Query Match 4.0%; Score 533; DB 4; Length 3080;
Best Local Similarity 18.6%; Pred. No. 1.6e-20;
Matches 561; Conservative 385; Mismatches 1041; Indels 1022; Gaps 138;

QY 29 QIARTHTDVL---LEVQHHSRDYASH-----LSPGSI 59
DB 465 QKRRHRRKGRDRHHRRHHKHADEGEDNNHNSNSNTNSISNNV 524

QY 60 QPQRRLSLLSEFPQNGERSQELHLRPE---SHSYLPELGKS---EMBFIESKRPLEL 112
DB 525 QPKRRR-----SGQOQHQAETETEPSTLSPEDKENFCRLQLLDNTNSAAIAV 574

QY 113 -----LPDPLLRPS-----PLLATQOP-----AGEDUTKOR 139
DB 575 ATSIADLSLSPVAHESETEQDEQVQPAVVEAQPPALTSAMAPAAANVKLPGRQTNNLS 634

QY 140 SLTGKL-----EPVSPSPHTDPELEVPPLSKKEELION----- 175
DB 635 SLTGNIFPGSAQOQOQSQEPSPSSSSQSGTTTASTVRKKQKAVILNGSGSSSTSK 694

QY 176 -MDRVREITWVQOISKLKKQOQLEBAAPPEPEKPVSPPIESKHSRLVQIIYDEN 234
DB 695 MADKADRNLSKTE---TTGSKSAAGSSAS-----SS 724

QY 235 RKKAEEAHRILEGLGPOVELPLYNQPSDTRQYHENIKINQAMKLLILYFKRRNHARKOW 294
DB 725 SSKAER-----ISG-----AYSXSGD-DFTETEEVLQIGHKVLVYVK--NH-RDAW 768

QY 295 -----KQFCQRY-----POLMEALEKKVIERIENPRRAKSKVREYVE 334
DB 769 PFVDPVEEDIAPRYIIIRPMDLLKWKEDKLDSEYHKTSEFRNDPRLIVNCRLYNGHN 828

QY 335 KOPPEIRKQELQRMQSRVQGRGSLMSAARSEHVESEI IDLGSEONLEKQMQOLAV 394
DB 829 NEYTEM--VNNLOQAF-----EKATKYFNLSDDEDD----- 860

QY 395 IPPMLYDADQORIKFINNGLMADPMKVYKDRQVMNMWSEQEKEETPREKFMQHPKNGFLI 454
DB 861 -PNLSYPAADSK-----MNVPEK-----YFSKKAKEETKDAQGRP-----AV 898

QY 455 ASFLEKRTVAECVLYYLTKQENYKSLVRSYRRRGKSQOQOQOQOQOQOQOQOQOQMP- 513
DB 899 SAAEEDLSEIEAPOAKQKRKEKED-KRRKKTKSKADVETDDDEMEAREPTPPPPP 957

QY 514 ---RSSQEEKDEKEKEKEKEKEKEPEVNDKEDL-----LKEKTP----- 551
DB 958 PTKSGKSKTKGKEKEKEKEKED-KDEKDEKTSFKGRKTKSKASKSKKTKYKGA 1016

QY 552 -----DTSGEDNDEKEA-----VASKGRKNTANSQGRK- 579
DB 1017 KXSADSDPESDPSDESSEDYDDHISLAKTKSLVKPTARTIAAKKKSVPAESKVXM 1076

QY 580 -----GRITR-----SWANEANSEEAITPOQSABELA---SMELNE---SSRW 615
DB 1077 PTPVKRQVKGKGGRKAKDDSLDSQSDVNVKQLPPTAAAALAEAAELEEDDDPPP 1136

QY 616 TEEMETAKGLLEHGRNWSAIARMVGSKTVSOCK-NFYFNKK---RQNLDEL----- 666
DB 1137 DEDEDSGR-----SRSMSPFKVDLHKYKSKSALNDLSELLTVVK 1179

QY 667 -----QHKLMKEKERNARR-----KKKAPAAAEAAFPVW----- 700
DB 1180 VPTAETTKLSARHODEADEERSRESGDPKSLNSRGSSEER--FPVAKKGKKAESSKK 1237

QY 701 -----EDEMEASGVSGNEEE--MVBEAEALHAGSNEV----- 731
DB 1238 EKEKGRDKDRDRDKEDKDKSRSAKDKKSKDSPLVAAAAEAAVVOAELMLLPMDK 1297

QY 732 -----PRGCSGSPATVNNSSDTSIPSPHTEAAKDTGQNGPKPPATLGADGPPGPP 783
DB 1298 YDVIKYRRSRAAFSGSSASNSLAPSDSKSAITKSNRENKASAKREKSPDAVENKGRK 1357

QY 784 TPRRTSRAPIEPTPASEATGAPTPPPAPPS--PSAPPPVPVPEEKEEETAAAPPVEGE 841
DB 1358 SKDQKRKSKESDKAEDKASKADTEKHSEKSKKEEPPKVKVKAQREKSPAPVSALETIK 1417

QY 842 EQKPPAAABELAVDTGKAEPKPVKSECTEEAEAGPAKGDAAEABATAEGALKAKKK--EGG 899
DB 1418 E--PP-----APTIAATSAGKVEKPAKPEKPKRDKQPPKPADKSEK 1464

QY 900 SGRATTAK-----SSGAPOSDSSATCSADEVDEAGGDKNRLSPSPSLTPTGDPANA 955
DB 1465 SGKSSSKAAQKAGQPTNNNTNLEALDVETEQLTKDINRWLEHTPRF---TEFSSASN 1521

QY 956 SPQK-----PLDLKQLKORAAAIIPIQVTKVHEPPREDAATKPPAPPAPPPP 1002
DB 1522 SPSRYNLLDDPDSGIGSKLDAADFRFPAL-----AAPKAELVPTKLA----- 1564

QY 1003 ONLQPSDAPQOPGSSPRGKSRSPAPPADKEAPAAEAQKLPGDPP----- 1047
DB 1565 EALNELVSEPKAVSKSESESVAPT-----SVSSCGTPPHSMHSGNSIGTSAT 1616

QY 1048 -----CMTSGLPFPVPPREVIVIKASPHADPSAFSYAPPCHPLPLGLHDTAREVLP 1101
DB 1617 AASSNCSNNSMPTLP---VAVTP-TPTPA---PPLLPIPKPEKPSPTTOLLNPP- 1666

QY 1102 ISNPPP-----LISSAKHPSVLERIGAISQMSVOLHVPYSEHAKAPVGVPTMGLPLM 1156
DB 1667 ---PPHHIKQOLAKEAKRSLKQAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAA 1708

QY 1157 DPKLAPFSGVQEQLSRPGQAGPPESLGVPTPAQEAASVLRGTAALGVSGSITKGPSTR 1216
DB 1709 -----DRLQP-GKA-----KGNLLQNV-----VAVKSTE 1731

QY 1217 VPSDSAITYRGSITHTGPADVLYKGTITRIIGEDSPSRIDRG-----REDSLPKGVHIYE 1271
DB 1732 EGGDS---HAGVNPVTIVKVKELKNALITETC-EGAP-KLSLGTIVLTKTQDFSLGKLEMS 1786

QY 1272 GKKGHVLSEYEGMSVTCSEKEDGRSSGPPHETAAPKRYDYMMEGRVGRASIASIEGLM 1331
DB 1787 GKGD-----ANEDDR-----PNEASPKNS----- 1806

QY 1332 GRAIPPERHSPHLLKEQHHI--RGSITQIGIPRSYVEAQEDYLREAKLLKRECTPPPPPP 1389
DB 1807 SPPTTNTTEAPKPFALHELKSKGKSSK---PSKSEASQKEKPNLSAWLKAFCGPKVSKK 1863

QY 1390 SRDLTEAYKT---QALGPLKL-KPAHE---GLVATVKE---AGRSIHEIPREELRH 1435
DB 1864 SED-EKQQTVPQDLOGDSKVAPPAHSAGDNFSLPTVMRQRPKPTGTSINSE--RSFSQ 1920

QY 1436 TPELPIAPRLKEGSIQTQ---TPLKYDTGASTTGS-----KKHDSRLSISPORTFP 1485
DB 1921 DPDSPRIAIDERYGVAAGSYTSPi---GASPIGASPIWVSPKPNDDMGKPAKP--- 1971

QY 1486 PVHPLDVMADALERACVYEE-SLKSRRPGTASSGSGSIARGAPVIVPELCKPQSPLYE 1544
DB 1972 --YPLN---GAIKVGFYQDTTKSSPKDKSCS-----PREMNSPY 2006

QY 1545 DHGAPFAGHL-----PRGSPVTMREPTPRLQBSL---SSSKASQDKLKTST----- 1588
DB 2007 Q-----YSQHIYSSASSPNVSTPDMSGTSPYGGNSNPNPSEASKTPAYSSTSP 2062

QY 1589 ---PREIAKSPHSTVPEHHPHFIPSPYEHLLRGVGVLDLYRSHIPLAFDPTSP 1645
DB 2063 YKQPSQESDYNSSMSPSTPNPHSPYQ-----QPOSSPYTTTPOQS 2102

QY 1646 AAAYLPRHLAPNPTYPHYLPYPPYRGYDPTAALENRQTIINDYITISQOHHNTATAMAQ 1705
DB 2103 QSTHPSPYH-NOSPVHQOQHSFY---HP-PAAQOQQQ-----SSQFSPHSAHQAQA- 2148

Db 2153 VATSTTPPAATILPKGPP---APATATP-----APTSPFSAATAGSMYSLVAPKAQRSP 2205
QY 2225 VSP-----PEGWTEPGHS-----RSNAVPLLYRDGEOTEPSRMGSKSPGNTSQ 2267
Db 2206 KAPQKVKAIAISIPVGSFAGASGRPGAPROPLEP-----GPFVREPTAPESELEGQPT- 2259
QY 2268 PPAFFSKLTESAMVSKKQEKINLKNTHNRNEPEYNISQPTETFFNMPAITGTLMTY 2327
Db 2260 PPA-----PPPIPETWTPA 2274
QY 2328 RSQAVQEHASTNMGLEAIRKALMGKYDQWESPPLSANAFNPLNASA--SLPAAMPITPA 2385
Db 2275 RS-----SPPLPPPAERTSAKGPETWASKPPSS 2304
QY 2386 ADGRSDHITLTPGGGKAKVSGRPSSRKAKSPAGLASDRPPSVSSVHSEGDCHRTPL 2445
Db 2305 SDWR-----VPGQGLENR--GEPTPPSPAPAPAVAGGSES--SSGRAAGD-----TP- 2350
QY 2446 TNRVWEDRPSSAGSTPPFYNPLIMRLQAGVMASPPP 2481
Db 2351 -----ERKEAGT-----GKKVKVRPPP 2368

RESULT 22

AAU33195
ID AAU33195 standard; protein; 2153 AA.
XX AAU33195;
XX AAU33195;
DT 18-DEC-2001 (first entry)
DX Novel human secreted protein #3686.
DE Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX Homo sapiens.
XX
XX WO200179449-A2.
XX
PD 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US008656.
XX
XX 18-APR-2000; 2000US-00552929.
XX 26-JAN-2001; 2001US-00770160.
XX (HYSSE-) HYSSEQ INC.

XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-611725/70.
DR

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.

XX Claim 20; Page 727-728; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon

CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX Sequence 2153 AA;

Query Match 4.0%; Score 522.5; DB 4; Length 2153;
Best Local Similarity 21.0%; Pred. No. 3.9e-20;
Matches 488; Conservative 213; Mismatches 809; Indels 813; Gaps 114;

QY 474 KKNENYKSLVRRSYRRRKSKSQQQQQQQQQQ-----QQQQQQQPMRPSQBEKEKEKEAE 530
Db 70 KGNDPNVSLVPKDGWGASKQEQSDPKSKSDASTAQPPESQPLPAS-----QT PAS 119
QY 531 KEEEPYEVENDKEDLLKEKTDGDDNDKDEKAVASKGRKTANSQGRKRGRITRSMANE- 589
Db 120 NQPKPPAAPENTPLV-----PSGV---KSWAQASVTHGAHGDGGRASLLSRFSREEF 170
QY 590 -----ANSEEAITPQGAELA-----SMELNESSRWTE-----EEMETAKKGLLEHGR 632
Db 171 PTLQAAGDQDKAAKERESAQSSGPGSLRPQNSITWRDGGGRGPDLE- GPD SKLHHGH 229
QY 633 N-----WSAIARMVSGK 644
Db 230 DPRGGLQSPGPPQFPFPRYRGMMPFMYPLPPPPYQGPYRYPTPDGFSRFPVAGPR 289
QY 645 TVSQCKNFYF-----NYKKRQNLDELLOHKLKMEKERNARKKKKAPAAASEEAFFPV 699
Db 290 GSGPPMLVFPVGRPSILKEDNLKFDQ-----LDQEND-----DGWAGAEHEVDYTBK 338
QY 700 VE--DEMEASGVSGNEEMVVEEAEALHASGNEVPRGECGSPATVNNSSDTSIESIPHTE 757
Db 339 LXFSEDEE---DGRDSEGAEGHRDSQSASGERP-PEADGKK--GNSPSEF-PTPKTA 391
QY 758 AAK-----DTQONGPKXPATLG-----ADGPPGPPPTPP-----RRTSRAPIE-- 795
Db 392 WAETSRPTEPETEGPPAPKPPPLPGDYPRGGGPKCPAPEDDEDAWRORRKKQSSEISLA 451
QY 796 -----PTPASEATGAPTPPPAPPSPSAPPV 821
Db 452 VERARRRREBERRMOEERRAAEKLRLDKFGAPDKRLKAEPAAPAAESTAPPDA 511
QY 822 VPKEKEBETAAAPVVEEGEOKPPAAEBELAVDTGKAEFPVKSECTEEAEEGPAKGDAB 881
Db 512 VPKE-----LPAPPA-----PP-----PASAPTPETEPEPAQAAPPAQ 544
QY 882 AAEATAEGALKAEKKEGGSGRATTAKSGAPODSSSATCSADEVD-----EAEAGDKNRL 937
Db 545 STFTP-----GVAAPTLVSGGGSSTSTSGSFEASPVPEQLPSKEGPEPPEE 592
QY 938 LSPRPS-----LLTPTGD-----PRANASPKPLDLKOLKQ-- 968
Db 593 VPPPTTPPVKVEPKGDGIGTRQPPSGGLGYPKQKSLPFRPQOQQEQQL-LKQOQQHQ 651
QY 969 -----RAAAIPTQVTKVHEPPREDAAPTAKPAPPAPPQNLQPPESDAPQOQSSPRGKSR 1024
Db 652 WQHQHOGSGAPPTPVPP--SPQPQVTLGAVPAQAAPPPP-----PKALYPGALGRPP----- 700
QY 1025 SPAPPADKEAF-----AAEAQKLPDPPC--WTSLG-LPFPVPPRE-----VIKASP--- 1067
Db 701 -PMPPMNFDPMMWMIIPYVVDRLQGRPPLEFYPFGVHPHGLVPRERSDLSLSEFPDR 759
QY 1068 HAPDFSAFSAVAPPGHPLPLGLHD--TARVPLRP-----PTISNP 1105
Db 760 HAPAMLRERTPPVDPKLAWGDVTTATPAERPLTSPLRQAADDDKMRSETFPVPP 819
QY 1106 PPLISSAKHPSVLERIQIGAISQGMVQLHVPYSEHAKAPVGVPTMGLPLP----- 1155
Db 820 PPYLAS--YPGPEN--GAPGPPIS--RFPLEEPGPRP-----LPWPPGSEVAKTIQ 865
QY 1156 -MDPKKLAPFSGVKQE--QL-----SPRGAQPPESLGVTQAQASVLRGTALGS 1202

Db 866 TPKPKKEPP-----KETAQITGDEAGKRLPASRGAGPP-----PPRESRT--ETRWCP 914
QY 1203 VPGSITKIGPSPRVPDSAITVR--GSI-----THGTADVLTKGTITRIIGEDSP 1252
Db 915 RFGSS--RRGIP---PEECAPRRAGPIKPPPTKVEELPKPLEQGDETPKPKPDP 969
QY 1253 SRLDRGR---EDSLPKGHVIEGKKGHVLSY-----1281
Db 970 LKITKGLGPKETPPNGNLSAPARLRDYSYERVGFTSCRGGRGEYFARGGRFGRTYG 1029
QY 1282 ----GMSVYQCSKE---DGR--SSGPPHETAAPKRTYDMMEGRAISSASIEGLWG 1332
Db 1030 GRCGRGRSEFRSYRFRGDDGGGTTGGNHPAPR-----GRHASETRSEGEY 1079
QY 1333 RAIP-----PERH---SPHLKEQHIRGSIQO-----GIPRSYVEAQED 1369
Db 1080 BEIPKRCRQSGTSETSHESDLAPDKEAPTKEGTLIQVPLAPPPPGAPPAPART 1139
QY 1370 YLRREAKLLKREGTP-----PPP-----PPSRDLTEAYKTOALGLKLKPAHEG 1413
Db 1140 -CRGGRVFTPRGVPSSRRGRGGRRPPQVCPGMSPPAKSL--APKKPPTGP--LPSSKEP 1194
QY 1414 LVATVKEAGRSIHEIPRELRHTPELPLA-----PRPKESITQGT 1456
Db 1195 L-----KEKLIPGLSPFVARGGNGSGNVGMEDGERPRRRRHGRAQQDK 1239
QY 1457 ----LKYDTGASTGSKKHDVRSLLIGSPRTFP-----PVHPLDMADARALERAC-- 1503
Db 1240 PPRFRLKQERENARGSE-----GXPSLTLPASAGPEALITVIVAPPPAAAK 1291
QY 1504 ----YEESLKSRPGTASSGSGSIARGAPVIVPE-LGKP-----1536
Db 1292 SPDLNSQNSQANEWETASESDFTSERRGDXEAPPPVLLTKAVGTGGGGGAVGI 1351
QY 1537 ----RQSPLYEDHGAFAGHLPRGSPVTWRETPRLQEGSLSSKAS-----1580
Db 1352 SAMSRLDLSORAKDLKRS-----FSSQRP--GMEQRNRRPGGKAGSSGSSGGGGGPG 1406
QY 1581 ----QDKLTSTPREIAKSPHST--VPEHHPHPISPYEHLRGV-----SGVDLYR 1625
Db 1407 GRTGPRGDKRSPKNSRPRPEEPGLPLPPPPSSSAVFRLOQVTHSNPAGTQQAL 1466
QY 1626 SHIPLAFDPTSIPIGPILOAAAAAYLPHLAPNTYPHYLYPYLIRGYPDYTALENRQTI 1685
Db 1467 AQLSSRQGSVTAAGGHRHKPGPPQAPQ--GPSPRPETRYEPQ-----1508
QY 1686 INDYITSQMHNTATAMAQRADMLRGL--SPRESSALNAAAGPRGIDLSOVPHLPVL 1743
Db 1509 VNSGLSSD-----PHFEPGPMVRGVGTGTPRDSAGVSPFPKRR-----ERPPRKPEL 1556
QY 1744 V-----PPTCGTGA-----TAMDRLAYLPTAPOP 1767
Db 1557 LQESLPPPHSSGFLGSKPGPGQPAESRDTGTTEALTPHWNRLHATSKSRVPTSMEP 1616
QY 1768 FSSRHSSSPSPGCPHTLTKPTTTSSESRERDRDREREKSIILSTTTTVEHAPIW 1827
Db 1617 W-----MEPLSP-----FEDVAGTEMSQSDSGVD-----LSGDSQVSSGPC- 1652
QY 1828 RPTQEGSGSGSGSGSGSGSPASHAHQHSPISTPTQDALQORPSVLHNTGMKGII 1887
Db 1653 ---SORSSPDGGLKGAEGPKRPGGSSPLNAVPCGPGSEPPRPPAPHD--GDRKEL 1708
QY 1888 TAVEPSKPTVL---RSTSTS-----SPVRPA-ATFPPA-----TH 1918
Db 1709 PREQPLPPGPIGTERSORTDRGTGPIRSHRPGPPVQGTSDKSDLRLVVGDSLKAE 1768
QY 1919 CPFGGTLGDIYVPLMEVLLPKBAPRVARPERPRADTGHAFKAPKPARSGLE-----1970
Db 1769 KELTASVTEAIPVSRDOWELLPSAAS-AEPQKNLDSGHC--VPEFSSSGQRLYPEVFG 1825
QY 1971 ---PASPSKSGSEPRPLVPVSGHATARTAKNLAPHAS-----PDPPAPASASDPHR 2023
Db 1826 SAGFSSSQISGGAMDSQLHNSGG---FRPGTPSLHPRSQPLYLPPGPAPPASALLSGVA 1882

QY 2024 EKTQSKPFS-IQELELRSLGYHSGSSYSPBGPVSPVSPSLTHDKGLPKHLEELDKSHL 2082
Db 1883 LKQQLDFSTMQATGELKLPAGGVLYPPPSF-LYSPAFCPSPLPDTSLLQVRQDL--PSPS 1940
QY 2083 EGELRPKOPCPVKLGGEAHLPHLR-----LPESQSSSP--L 2119
Db 1941 DFYSTPLQ-----GGSGFLPSGAPAOQMLLPVMSQLPVNVFGLPPAPPPLSL 1995
QY 2120 LQTAFCVKGHQVAVTLAQHISEVITQDYTRHHP-----QQLSAPL- 2159
Db 1996 LPVGFALQPPSFVVRPQSSPSTGVLPXLARPPVYFGRTELHPVNKPFDFQKLSNLG 2055
QY 2160 -----PAPLYSPGASCPVLDLR-----RPPSDLY--LPPPD- 2189
Db 2056 GPGSSRTPTGRRPSSRLRSFGLNSRLQSLNSLTSGVFRNQASTFYQAGLPHPDALR 2115
QY 2190 -----HCAPARGSPHSEGGKRSPEPNKTSVLGGGEGIEP 2224
Db 2116 WIPKPMWERTGRPRDGP-----SRAEPPGSR---GDKEPGLPP 2151

RESULT 23

ABB65480
ID ABB65480 standard; protein; 3536 AA.

XX ABB65480;

XX AC AC

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 23232.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX Drosophila; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09583.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions.

XX Disclosure; SEQ ID NO 23232; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

XX ABB72072). The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at fcp.wipo.int/pub/published_pct_sequences

XX Sequence 3536 AA;

Query Match 3.9%; Score 519; DB 4; Length 3536;

Best Local Similarity 19.0%; Pred. No. 1.2e-19; Matches 589; Conservative 364; Mismatches 1037; Indels 1112; Gaps 136;	
QY 42 YOHSHRDYASHLSPGSIOPORRRPSSLSEFOP-----GNERSO-ELHLRPESHVYLP 95	Db 1076 APNVSVHETE-PASPTPSKQL--MPTPTTTPAPATKSPAA-----AAVAPFCAAN 1124
Db 199 FOHLSQHADY---FACQPCRLR-----FOADHLLGHQOQHSHHPEDGHATKPL 249	QY 842 EQKPP-----AAELAVDTGKAEPVKSECT-----EAELEGPAKDAEAAEAT 886
QY 96 -GKSE-----MEFIESKRP-----RLLELLDPLLRPSPLLATGQAPAGSEDLTKDRSLT 142	Db 1125 RQTTNSTVKSRSKRLSDCIAMLTGKLEKLTQTEVPPGPGLOKEKEKEQQDPAKQDKT 1184
Db 250 MGADEPVLFRGLGTQNLFSHRNRNQKEEPPPLATPS-----SKSRSSS 295	QY 887 AE-----GALKAKEGSGRATTAKSGAPO-----DSOSSATCSADEVDEAGGDKN 935
QY 143 GKLEVPSPSPHTDELELVPRLSKEBLIQMDRVREITWVEQOISKLKKKQOOLEE 202	Db 1185 PQVDHQSPAAAVAEKER---VQPKTKRAVSRRIIKMDATPETTAEPIQER----- 1235
Db 296 -----RAATORQOONHOHQOSQSHOQ 317	QY 936 RLLSRPRLTPTGDPRA-----NASQPKLDLKLQKQRAAAIPIQVTKVHEPPREDA 989
QY 203 EAAKPEP-EKPVSPPI-----ESKHSVLQIYYDENRK-----AEAHRILEGLG 249	Db 1236 -----KPVAMPAPVPPVAPAILNAAPAPALQPIIFASMPVAVVVPVAPVPPVAAA 1290
Db 318 TAAAGMALQESVNDPAVFRLLPASMVSSASVANSVDFDKFYKDVINVRHNLQNHLD 377	QY 990 APTK-----PAP-----PQNLQPSDAPQOQSSPRGSRSPAPADKE 1033
QY 250 PQV-----ELPLYNQPSDT-----ROYHENIKINQAMRKKLIL--- 282	Db 1291 ADLPVAIVPLVPAAPAPPVATALVHPTRTPTKYAAAKQMGAPP-----PKPPASLA 1344
Db 378 GRLEFTSSSTTSATSSASHKSDLIMLHGPSQPGGASYPTLLTADQFGGTGELLPLA 437	QY 1034 AFABAQKLPDPPCWTSGLPVPPPREVIKASPHA-----PDPSAFSYAPPGHP 1083
QY 283 YKGRNHAKQWKQFCQYDQIMEALEKKVERIENPRRAKESKVRYYEKGQPEIRK 342	Db 1345 A-----LKQYPPLEAT---LPVPTNSAPPVAVPLVPPVPPVPMPLGMQGW 1394
Db 438 KFRRRPHTKHSWKW-----DYVKKFTLINE 464	QY 1084 LPLGLHDTARVLP-----RPPTISNPPP-----LISSAKHPSVLERQIGALSO--- 1127
QY 343 QRELQERMO--SRVGORG-SGLSM---SAARSEHEVSEIIDGLSBOENLEKO--MRQAV 394	Db 1395 VEVNYN--MLPKLPLYMTQLQAPANPNPNPAPTAMFVMDHQPLNLTQSRGVLQKPLJ 1452
Db 465 GGRLVKLLQPNYGLURDLSKJDMWTQLTWRQKHJLSL---SLTBEQNEQORRLLEQLNL 521	QY 1128 -GMSVQLHVPYSEHAKAPVGPV-----TWGLPLPMDPKKLAPFSGVKQEQOLSP 1174
QY 395 I-----PMLYDADQOR-IKENM--NGLMADP-----MKVYKDRQVMNMW----- 432	Db 1453 SGTGDPGLVGLQHRAGMFPARROTICGFERNLIGDMEMEPDLDSKSKSRKSPVPP 1512
Db 522 ILDHRLPHIILEQNEQAVIKCENBEDEGLDHPGDDLPTQTFADESFLSLLOLQPRGG 581	QY 1175 --RGQAGPESIGVPTAQEASVLRGTALGSPVG-----SITKGIPS 1214
QY 433 -SEQEKETREKPMQ-----HPKNFGLI---ASFLEKTVAR----- 465	Db 1513 PRMQQELMPLVPLPLVTNANAL---APQSPGGAPLPTPVGGAPVTCQLPMSGVSVBG 1569
Db 582 SSEQERQERDQERSNRQKAVVLSGEWARPRLYLCICCGAKFDQKSLSEHKTFRHSI 641	QY 1215 TRVPSDSALTIRGSIHTGTPADVLYKGTITRIIGDSDSPRLDRGSDSLPKGHVIEGKK 1274
QY 466 CVLYYVLTKNENYSLVRSY---RRGK-----SQOQOQOQOQOQ 504	Db 1570 --VPAPLASHYYSNLDLKIIPQVRNPQGTG--VVPSSGPPAVSAASAAPHVPKSSG-- 1624
Db 642 YATHYEVVRELLAGNLLRHLFIPKRALGRFAAASCIRWPOIAPATVQOPKQVOPE 701	QY 1275 GHVLSYEGGMSVTQCSKEDGRSSGPPHETAAKRTYDMEGRVGRTASSAIEGLMGR 1334
QY 505 QO-----QOQPMRPS----- 515	Db 1625 GKRTTEG-----TGSKKD-----NGKQAKACP-RMDEVNHHIDDALNSVIM--AVQAS 1673
Db 702 EETRASSASSRSEVSTPTPLSALEQOQSPASGSGNASSSSSCSPSSASTLLSTATR 761	QY 1335 IPPERHSHPHLKEQHHRGHSITQGIPIRSYVEAQEDYLREAKLLKREGTPPPPPSRDIT 1394
QY 516 -----SQEEK 520	Db 1674 PDDDEERQAKKE-----KEREREREREKELEEPPLKFSNCI- 1714
Db 762 TSMNSTCTSSAATTVTSSTSPSSCTKGRKCSGLMDLYRHMDCSDYVWSLAKKK 821	QY 1395 EAYKTQALGPLKLKAHEGLVATVKEAGRSIHEIPRELRHTPELP--LAPPLKEGSGIT 1452
QY 521 -----DEKEKEKEKEKEE-KPEVENDKEDLLKEKTDDTSGEDNDEKEAVAS 566	Db 1715 -----MPLQCAEVLPSAVLEKSQAVNAIPAIVAPPVAVVVPVLAPOQVPVKNL- 1763
Db 822 YRYCCTKRRAFSKLQLSARKKEKLEVEEGTGDGDAEVDGDGEAEAGEESVS 881	QY 1453 QGTPLKYDTGASTTCKKHVRSLLGS-----PGRTFPPVHPLDVADARALERACYEES 1507
QY 567 KGRKTANSQGRKRG---RITRSMAN-----EANGEEAITPOQSAELASMELN 610	Db 1764 --TPKK-----RSMRSRTIDCSALLALEETLPAASLPCQICPLPVGDAK-----EE 1808
Db 882 SSKLKNRQRPQSDAESIRKLENLPAKRI CKKIFPVDNKAKRKAKNKAKAKSMAKSIV 941	QY 1508 LKSREGTSSSGSARGAPVIV-PELGKPRQSPITYEDHGAPFAGHLPRGSPVTMREPT 1566
QY 611 ESSRWTEEMETAKGLLEH-----GRNWSAIARMVGSKTSQCKNFYFNKKQKNLDE 664	Db 1809 MBSLP-----DPIAPTAPVVEPOL-----PVQVQSGT 1836
Db 942 KSKSKSKQORSSTKRIYNQHLRTTRSRGRSSVATGSSAAAAVAQQQSRKQKQOQK 1001	QY 1567 PRLOGSSLSKSKASQDRKLTSTPREIAKSPHSTVEHHHPHPIPSYEHLLRGSVGDLYRS 1626
QY 665 ILQOHLKWEKERNARRKKKAPAAASEEAPPPVVEDEMEASGVSGNEEEMVEAEAL 724	Db 1837 TSLRDKKTETVMPCNTSTTIPTATPLAESPE-----PDVP-----V 1874
Db 1002 QOQOQOQOQOQAKLEKVGKPPPEP-----PAAVESQMPA----- 1038	QY 1627 HIPLAFDPTSPRGIPLO-----AAAAYLPRHLAPNPTVPHLYPPYLIRGYPDTAALENR 1682
QY 725 HASGNEVPREGSGPA---TVNNSSTESIPSPHTEAANDTGQNGPKPPATLGCADGPPPG 781	Db 1875 NLPOA--ETSVPAPVPPVSVISVAPAVLPLPTRTATPP-----PTTMAETNC 1918
Db 1039 -----KTEARAPATVTAKSRSAPAEKPSLRLEL-----PQTL-----PEAA 1075	QY 1683 QTIINDYTTSQOMHNTATAMAQADMLRGLSPRESSALNYAAGPRGIIDLSQVPHLPV 1742
QY 782 PTPPPRTSRAPLEPTPASEATGATPPPPAPSPSPAPPPVVPKKEEBETAAPPVEEGE 841	Db 1919 SSLMEEH--SSNLNNNTSS-----GPHSLAQ-SEQPI 1947


```
Db 2558 SEVLNMQSEYRLHPYTPVRDVRIMVHPHTAVSEQPR-----AAGGVKVPBPASKA 2608
QY 2232 TEPGHSRAVYPLLYLDEGEQTEPSRGMGSKSPGNTSQP 2268
Db 2609 PQ-----QFGKEAAKTPDAAKAP 2626

RESULT 25
ABB33490
ID ABB33490 standard; peptide; 2665 AA.
XX ABB33490;
AC ABB33490;
DT 04-FEB-2002 (first entry)
XX
DE Peptide #996 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
XX
OS WO200157277-A2.
XX
PN 09-AUG-2001.
XX
PD 30-JAN-2001; 2001WO-US000669.
XX
PF 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human foetal liver.
XX
PS Claim 27; SEQ ID NO 26125; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: the sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2665 AA;

Query Match
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

QY 71 EFQPGNERSQELHLRPE-----SHSLPELGKS-----EMEFIESKR---PRLELLP 114
Db 525 ERKSGQKSHSVNTEEKIGIDIDHTQSYRQKQEQRRKQKQOMEIAKSKFGSPKKDV-- 582
QY 115 DPLLRSPLLATQOPAGSGLTKDRSLTGKLEVPSPSPPTDPELELVPPRLSKLELIQ 174
Db 583 DEYERSLVHEVGKP--PQDVTD-----SPFSKK-----X 611
QY 175 NMDRVDEI--TWVQOISLKKKKQOOLEBEAAKPP-----EPEKVPSPPIES----- 221
Db 612 RMDHVDICTKERNYRSRQISEDSERTGGSPSVRHGSFHEDEPDIGSPRLLSVKGSP 671
```

```
QY 222 -----KHSRLVQIYYDENRKK--AEAHRILEGLGPOVELPLYNQPSDTRQ 265
Db 672 KYDEKVLPSYTNITVREESLKFNPYDSRRQEMADMAKIKLSVLNSDELNRW-----DSQM 727
QY 266 YHENIKINQAMRKKLILYFKRRNHARKQWKFCQBYDQIMEALEKKVRIENPPR--- 322
Db 728 KQDAGRFVSPNSII---KDSLRKRSVRDL--EPGEVPSDSDESHKSHSPASAL 781
QY 323 -----RAKESKVREYVEKQPEIRKRELQERMOSQVQQRGSGI-----SMSAAR 367
Db 782 YESSRLSFLLRDRDKLRDERLSSLSERNFYSPALDKTITPTKALLERAKSLSSSR 841
QY 368 SEHEVSEIIDGLSE-----QENLEKQMRQLAVIPPLYDADQOQRIKFINNGLMADPMKY 423
Db 842 EEN--WSFLDWSRFANFRNNKDEKVDSPRPIPSWYMKKKIR-TDSEGMDDKKEH 898
QY 424 KQOVNMWSEOEKETFREKFMQHPKNFGLIASFLERKTVAECVLYYYLTKNENYKSLV 483
Db 899 KEE-----QERQELFASRFL-HSIFRQDSKRLQ-----HLERKEEDSDFIS 940
QY 484 RSYRRRGKSGQQQQQQQQQQQQQQQ-----MPSRQOEKDEKEKEKEKEE 533
Db 941 GRIY---GK---QTSSEGANSTTDSIQEPVVLFSRPMELTRMQCKEKEKQPKVEKOE 994
QY 534 EKPEVENDKEDLLKEKTDGSDNDEKEAVASKGRKTANSQGRKRGRITRSMANEASE 593
Db 995 ---DTEN-----HPKTPESAPEND-----SELKTP 1017
QY 594 EAITPQQSAELASMEINLESSRWTEBEMETAKGLLEHGRNWSAIAIRMGVSKTVSQCKNFY 653
Db 1018 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1046
QY 654 FNYKKRQNLDELQOHLKWEKERNARRKKKAPAAASEEAFPPVVEDEMEASGVSN 713
Db 1047 -----EAPLVTEKTVSEAPKSPAPAPA 1075
QY 714 EEMVEEAEALHASGNVPRGCSGPATVNNSSDTSIESPHTAAKDTGQNGPKPATL 773
Db 1076 PVEQLEQV-----DLPPGA-----DPDKEAAMP---AGVESSGSDQPP-YL 1114
QY 774 GADGPPGPPPTPRKTSRAP--IEPTPAEATGAPTPPPAPSPSPAPPVVPVKEKEEBET 831
Db 1115 DAK-----PPTGASFSQAESNVDPEDPS-----TQPLSKPAQKSEANEKPAKPDAT 1163
QY 832 AAAP-----VEEGEQKPPAAEELAVD---TGKAEPEVKSECTEBEAEAGPAKDAEAA 883
Db 1164 ADAEPDANOQAAAPESQPPASEDLEVDPPVAAKOKKPNKSKRSTFPVQAAAVSIVEKPV 1223
QY 884 EATAEGALKAEKKEGSGRATTAKSGAPQDSDS-SATCSADEVDEAEAGDKNRLLSPPR 942
Db 1224 TRKSERIDREKLKRSNPRGEAQKLELQWEAEKTRTASKNSAADLE-----HPBP 1275
QY 943 SL-LTPT-----GDPRANASQK-PLD-----LKOLKQAAAIPIPTQVTK 980
Db 1276 SLPLSKTRRRNRSVVATMGD-HENRSVKPEPVQPRVTRKRLERLEQEAAPV---TTPR 1332
QY 981 VHEPPR-----EDAAPTKPAAPP-----PONL-----QPSDAP 1012
Db 1333 RGRPPKTRRRARDEBENEAKPAETLKPEGWRSRPSQKTAAGGGQGGKKNPEKVDAT 1392
QY 1013 QQPGSSPRG-----KSRSPAPPADKEAPAAEAQKLPDGPCCWTSGLPFPVPREVIKASP 1067
Db 1393 REATEIVGQGVKVESMEPKAAEEAGSEKQRDKD--AGTDKNPPTAPVVEVVEKXP 1450
QY 1068 HAPDPSAFSYPAGHPPLPLGLHDTARPVLPFRPTTISNPPPLISSAKHPSVLQRIGAIQ 1127
Db 1451 -APEKNSK-----KEGRSRNSRLAVDKSASLKN-----VDAAVSPRGAQAQAGERES 1497
QY 1128 GMSVOLHPYPSHAKAPGVPTMGVPLPMDP-----KKLAPFSG-----VKQBL-- 1172
Db 1498 GV-VAVSPKESRSPQKEDGLSQLKSDPVDPDKEPEKEDVSASGSPSEATQAKOMELBQ 1556
```

QY 1173 -----SRGQAGPPESLGVPTAQEASVLRGTALGSV----- 1203
Db 1557 AVEHIAKLAESAASAAKADAEGLA--PEDRDKPAHQASQETELAAAGISINDISGEPE 1614
QY 1204 -----PGSITKGIPTTRV-----PSDSALTYRGSITHGTPADVLVYKGTI--TRIIGED 1250
Db 1615 NFPAPPYPGESQTDLOPPAGAAQAPSE-----EGMETDEAVSGILETAATES 1664
QY 1251 S-----PSRLDRGRDLSLPKHVIVYEGKKGHVLSEYEGGMSVTQCSEKDRSSS 1298
Db 1665 SRPPVNAPOPSAGPTDTKARGNSSETSHVPEAKGSK-----EVEVTIVRDKGRQ-- 1716
QY 1299 GPPHETAAPKRTYDMMEGRVGRASIASIEGLMGRAPPERHSPHKLKEQHIRG---SI 1355
Db 1717 -----KTRRSRRKRNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
QY 1356 TQIGIPRSYVBAQEDYLRRRAKLLKREGTTPPPPPPSRDLTEAYKTQALGPLKPAHEGLV 1415
Db 1753 NEGTTVQHPEAPQ-----BEKQSEKPHSTPOSCSTDLSKIPSTE----- 1792
QY 1416 ATVKEAGRSIHIEPRELRHT-----PELPLAPRPLKEGSIQTGTPKLYDTCGASTTGSKK 1470
Db 1793 -----NSQSELSVEERTPTKASVDPDLPPPPQ-----APVDEPOA-----R 1830
QY 1471 HDVRSLLIGSPGRTP--PVHPLDVMDARALERACYEESLKRPGTASSSGSIAAGAPV 1528
Db 1831 FRVHSIIESDPVTPPSDPSIPIPTLPSV-----TAAKLSPPVASG--- 1870
QY 1529 IYPELCKPRQSP-----LTVEDHGAAPFAGHLPRGSPVTMREPTPRLOEGSLSS---SK 1578
Db 1871 -----GIPHQSPPTKVTEMITRQE-----EPRAQSTPSPALPDPTKASDVDT 1913
QY 1579 ASODKRLTSTPREIAKS-----PHSTVPEHPHPISPYEHLRGVSGVDLYR 1625
Db 1914 SSTLRKILMDPKVTSATSVTSTVTTAIEPVSAAPELHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDPTSIPIRGIELDAAAYLPHRLAPNPTYPHLYPYLYNGYPTDAALENQRTI 1685
Db 1968 KTAPPVTNNSEIQASEVLVAADKEKVAPIAKIT-----SVISRMPPVSDLENSQKI 2020
QY 1686 INDYITSOQHHNTATAMAQADMLRGLSPRESSLALNVAAGPRGIIDLQVPHLPVLVP 1745
Db 2021 -----TLAKPAPQTLTGL-----VSALTGLVNVSLVP--VNALKG 2053
QY 1746 PTPGTATAMDRLAYLPTAPQPPSSRRSSPLSPGGPTHLTPTTTSSSERERDRDRD 1805
Db 2054 PVKGSVTTLLKSLVS-----TPAGPVNVLGPV----- 2080
QY 1806 RDREREKSILTSTTTVEHAPI-----WRPTEQSSGS-----SGSS 1841
Db 2081 -----NVLGTGVNVLTTPVNATVGTVNAAPGTVNAASAVNATASAVTVTAGAVTAAS 2133
QY 1842 GGGGGS-----SSRPASHAHQHSIPSPRTQDALQORPSPVLHNTGMGIIT 1888
Db 2134 GGVATATGTVTWAGAVIAPSTKCKQASANENSRFPGSMPIVDRPA---DAG--SGAGL 2189
QY 1889 AVEPSKPTVLRSTS---TSSVPKPA--TFPPATHCLPGTGLDGYVTLMEVLLPKPA 1942
Db 2190 RVNTSEGVLLSVGGKTEGPQISAKISQIIPAS-----AMDIEQOSVSKSVKQKPS 2243
QY 1943 PRVARE--BRPRADTGHAFIA-----KPPARSGLPASPSPKSGSEPRPL--- 1984
Db 2244 VTASQPSKGPQAPAGYANVATHSLTLVTAQTNASPVISVVK--ADRPSTL--EKPEFIHLS 2301
QY 1985 -----VPPVSGHATTARTPA----- 1999
Db 2302 VSTPVTGGGTVKVLTQGINTPPVLVNQLVTSIVTTNKKLADPVLTKIETKVLQPANL 2361
QY 2000 -KNLAPHASDPDPAPPASADP--HREKTQSKPF--SICELELRSLGVHSGSYSPGVE 2054
Db 2362 GSTLTTHH---PPALPSPKLPTVENVHVPSPSPADRTVSHLAAAKLDHSPSPGSPS 2417
QY 2055 PVSVPSSPSLTHDKGL-----PKH---LEELDKSHL-- 2082

Db 2418 SPFRASHFSSASTALSTNATVLAAGIPVQFIISSIHPEQSVIMPPHSITQTVSLSHLS 2477
QY 2083 EGELRPKOPG---PVKLGGEAAHLPHLRPLRPESQSSPSSPLLQTAGVKGHQVRVVTLAQH 2138
Db 2478 QGEVRMNTPTLPSIYYSIRPEALHSER-APL---OP-----QQIEVRA-- 2516
QY 2139 ISEVITQDYTRHHPOOLSAPLPAPLYSPFGASCPVLDLRRPSPD---LYLPPPDHGAAPAR 2195
Db 2517 -----PQRASTPQAP-----AGVPALASQHPPEEVRVHPLPVARATAPVQ 2557
QY 2196 GS-----PHSEGKRSPEPNKTSVLGGEDGIEPVPSPPEGM 2231
Db 2558 SEVLVMQSEYRLHPYTPRDRVIMVHPHTAVSEQPR-----AADGVVVKVPPASKA 2608
QY 2232 TEPGHSRSNAVVELLYRDGEQTEPSRMGSKSPGNTSOP 2268
Db 2609 PQ-----OPGKAAKTPDAKAP 2626
RESULT 26
AAM26950
ID AAM26950 standard; protein; 2665 AA.
XX AC AAM26950;
XX XX
DT 17-OCT-2001 (first entry)
XX DE Peptide #987 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS genetic disorder.
XX OS Homo sapiens.
XX XX
PN WO200157272-A2.
XX PD 09-AUG-2001.
XX XX
PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488997/53.
DR XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 27219; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs;
CC see AAI31315-AAI5746). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX SQ Sequence 2665 AA;
Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

Db 2244 VTASQPPSKGQAPAGYANVATHSTLVLTAQTVNASPVISVK-ADRPSEL-EKPEPIHLS 2301
QY 1985 -----VPPVSGHATARTPA----- 1999
Db 2302 VSTPVTQGGTVKVLTOGINTPPVLVHNQVLTPSIVTTNKKLADPVLTKIETKVLQPANL 2361
QY 2000 -KNLAPHASPPPPASADP--HREKTQSKPF--STQLELRSLGVHGSYSPEGVE 2054
Db 2362 GSTLTTEHH-----PPALPSKLTPEVNHVPSGSPIPADRTVSHLAAAKLDAHSRPSGPGPS 2417
QY 2055 PVSVPSSPLTHDKGL-----PKH--LEELDKSHL- 2082
Db 2418 SFRASHPSSTASTALSTWNLWLAGIPVPOFISIHPEQSVIMPPHSITQVLSHLUS 2477
QY 2083 EGELRPKQPG-----PVKLGEAAHLPLRLPESQSSPPLQTAGVKGHQRVVTLAGH 2138
Db 2478 QGEVMTTLPSITYSIRPEALHSR-APL---QP-----QQIEVRA-- 2516
QY 2139 ISEVITQDYTRHHQOOLSAPLPAPLYSFFGASCPLVLDLRPPSD---LYLPPPDHGAPAR 2195
Db 2517 -----PQASTPQAP-----AGVPALASQHPPEEVEVHPLVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGGEGIEPVSPPGCM 2231
Db 2558 SEVLVMQSEYRLHPYTVPRDVRIMVHPHTAVSEQPR-----AADGVVVKVPPASKA 2608
QY 2232 TPGHRSRVAVPLLYRDGQTEPSRMGSKSPGNTSOP 2268
Db 2609 PQ-----QPGEAAKTPDAKAP 2626

RESULT 27
ID ABB28314
XX ABB28314 standard; peptide; 2665 AA.
AC ABB28314;
XX
DT 01-FEB-2002 (first entry)
DE Human peptide #965 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer.
XX Homo sapiens.
XX WO200157271-A2.
PN
XX
PD 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
PS Claim 27; SEQ ID NO 11282; 327pp + Sequence Listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2665 AA;
Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

QY 71 EFQPGNRSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR-----PRLLELP 114
Db 525 ERKSGQEKSHSVNTEEKIGIDIDITQSYRKQMEQSRKQKQEMETAKSEKFGSPKDV-- 582
QY 115 DPLLRPSPLLATGQAGSEDLTKDRSLTKGLPEVSPPPPHDTPELELVPPRLSKELLIQ 174
Db 583 DEYERRSLVHEVGKP--PDVTDD-----SPPSKK-----K 611
QY 175 NMDRVDREI-TWVEQOIGLKKKQOOLEEAAKPP-----EPEKVPSPPPPIES----- 221
Db 612 RMDHVDFDICTKERNYRSSRQISDSERTGSPSVRHGSHFDEDPGSPLLSVKSGSP 671
QY 222 -----KHSRLVQIYDENRKK--AEAAHRIEGLGQVQLPELVPLYNQPSDTRQ 265
Db 672 KVDEKVLPSYNTVREESLKFNPYDSRRQEMADMAKIKLSVLNSEDENRW-----DSQM 727
QY 266 YHENIKINQAMKLLILYFKRNHARKQWKQFCQRYDQLMALAEKVKYRIENRR-- 322
Db 728 KQDAGRFDVSPNSII-----KDSLRKRSVRDL--EPGEVPSDSDEDEGHHKSHSPRASAL 781
QY 323 -----RAKESKVEYVEKQFPEIRKQRELQERQSRVQSGSGL-----SMSAAR 367
Db 782 YESSRLSFLLRDREKLRDERLSSSLERNKFYSFALDKTITPTKALLERAKSLSR 841
QY 368 SEHEVSEIIDGLSE-----QENLEKQMRQAVITPPMLYDADQORIKFINNGLMADPMKVY 423
Db 842 EEN--WSFLDMSRPFANFRNNKDEKVDSPRPIPSWYMKKKIR--TDSEGMDDKDEH 898
QY 424 KDRQVMNMSQEKETFREKMQHPKNECLIASFLERKTVAECVLYYLTKEENYKSLV 483
Db 899 KEEE-----QERQELFASRFL-HSSIPEQDSKRLQ-----HLERKEEDSDFTS 940
QY 484 RRSYRRRGKSQQOQQOQQOQQOQQOQQO-----MPRSSQBEKDEKEKEKEAEKEE 533
Db 941 GRIY--GK--QTSEGANSTTDSIQEPVVLPHSRFPMELTRMQQKEKQKQKPEKQKE 994
QY 534 EKPEVENDKEDLLKKTDDTSGEDNDEKEAVASKGRKTANSQGRKRGITRSMANEANSE 593
Db 995 ---DTEN-----HPKTPESAPENKD-----SELKTP 1017
QY 594 EAITPQQSNELASMEINLESSRWTESEMETAKGELLEHGRNWSAIARMVGSKTVSQCKNRY 653
Db 1018 PSVGPPSV-----TVVTLSSAP-----SALEKTTGDKTV----- 1046
QY 654 FNYKKRQNLDELQOHLKMEKERNARRKKKAPAAAGSEAAFPVVEDEEAEASVGN 713
Db 1047 -----EAPLVTEKTVETPATVSEAKPASEPAPA 1075
QY 714 EEMVEEAEALHASGNEVPRGECSPATVNNSSDSTESIPSPHTEAAKOTGQNGPKPPATL 773

PD 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000666.
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 DR Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts.
 XX Claim 15; SEQ ID NO 20720; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 2665 AA;
 SQ
 Query Match 3.9%; Score 518; DB 4; Length 2665;
 Best Local Similarity 19.3%; Pred. No. 9.2e-20;
 Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;
 QY 71 EFQPGNERSQELHLRPE-----SHSYLPELGS-----EMEFIESKR---PRLELPL 114
 DB 525 ERKSGQKSHSVNTEBKIGIDHTQSYRKQEQSRKQKQMEIAKSEKFGSPKDV-- 582
 QY 115 DPLLRSPLATGCPAGSEDLTKDRSLTCKLRFVSPSPPHDTDPLELVPPRLSKBELIQ 174
 DB 583 DEYERSLVHEVGK--PQDVTD-----SPSKK-----K 611
 QY 175 NMDRVDRDI-TWVEQOISKLKKQKQLEBEAAKPP-----EPEKVPSPPIES----- 221
 DB 612 RMDHVDFDICTRERNYRSRQISESETGSPSVRHSFHEDEPIGSPRLLSVKGSP 671
 QY 222 -----KHSRLVQIYDENRK--AEAAHRILEGPGQVELPLYNQPSDTRQ 265
 DB 672 KVDEKVLPSYNTVRESLKFNPYDSRREQADMAKIKLSVLNSEDENRW-----DSQM 727
 QY 266 YHENIKINQAMRKLILYFRNRHARKQWKFCQRYDOLMEALEKKVRIENPR--- 322
 DB 728 KQDAGRFDVSPFNSII-----KDSLRKRSVRDL--EPGEVPSDSDEDEGKHSFRASAL 781
 QY 323 -----RAKESKVRYEYKOPPEIRKQRELQRMQSRVQQRGSL-----SMSAAR 367
 DB 782 YSSRLSFLLRREDKLRDERLSSLSERNKFYSFALDKTTTPDKALLERAKLSLSR 841
 QY 368 SEHEVSEIIDGLSE-----QENLEKQMRQLAVIPPMYLDADQOQRIKFINMGLMADPMKY 423
 DB 842 EEN--WSFLDWDSRFANFRNNKDEKVDAPRIPSWYMKKKIR-TDSEGRKMDKDEH 898
 QY 424 KROQVNMNWEQEKETFRFKFQHPKNGFLIASFLERKTVASCVLYYYLTKNENYKSLV 483
 DB 899 KEE-----QBRQLFASRFL-HSIFQDSKRLQ-----HLERKEEDSDFIS 940
 QY 484 RRSYRRRGKSQQQQQQQQQQQQQQQQQ-----MPRSSQBEKDEKEKEKEKEE 533
 DB 941 GRIY---GK---QTEGANSITDTSIQEPVVLPHSRFEMELTRMQOKEKQDPKEVEKE 994
 QY 534 EKPEVENDKEDLLKKTDDTSGEDNDEKEAVASKORKTANSQORRKGRITRSMANEANSE 593
 DB 995 ---DTEN---HPKTPESAPENKD-----SELKTP 1017
 QY 594 EAITQQQSAELASMELNESRWTEEMETAKGGLLEHGRNWSAIARMVGSKTVSQCKNPFY 653
 DB 1018 PSVGPPSV-----TVVTLESAP-----SALEKTTGDKTV----- 1046
 QY 654 FNYKKQNLDEITLQHKLMKEKERNARRKKKAPAAAESEAAFPVPVVEDEEAEASGVS 713
 DB 1047 -----EAPLVTTEKTVEPATVSEAKPASEPAPA 1075
 QY 714 EEMVVEEAALHASGNEVPRGECSPATVNNSSDTSIPSPHTEAAKOTGQNGPKPPATL 773
 DB 1076 PVEQLEQV-----DLPPGA-----AGVEBEGSGDQPP-YL 1114
 QY 774 GADGPPGPPPTPPRTSRAP--IEPTPASEATGAPTPPAPSPAPSPAPPPVPVPEKEKEE 831
 DB 1115 DAK-----PPTFGASFQAESNVDPEDS-----TQPLSKPAQKSEANEPEAKPDAT 1163
 QY 832 AAAPP-----VEEGEQKPPAAAEELAVD---TGKAEPPVKSECTEEAEEGPAKGDAEAA 883
 DB 1164 ADAEPDANKAAEAPESQPPASEDLEVPVNAKKKPNKSKTPTVQAAAVSIVEKPV 1223
 QY 884 EATAEGALKAEKKEGGSGRATTAKSSGAPQSDS-SATCSADEVDEAEGDKNRLSPRP 942
 DB 1224 TRKSERIDREKLKRGNSPRGEAKQLLELMKEAKITRTASKNSAADLE-----HPBP 1275
 QY 943 SL-LTFT-----GDPRANASQK-PLD-----LKQLKQEAALPIPIQVTK 980
 DB 1276 SLPLSTRRRNRVSVYATMGD--HENRSPVKEPVEQPRVTRKRLERELAEAAVP--TTPR 1332
 QY 981 VHEPPR-----EDAAFTKPAAPPAPP-----PQNL-----QPESDAP 1012
 DB 1333 RGRPKTRRRADEEENEAKEPATLKPPEGWRSPRSQKTAAGGPGQKKGNEPKVDAT 1392
 QY 1013 QPGSSSPRG-----KSRSPAPPAQKEAFAAEAQKLPDGPWCWTSGLPLFPVPVPREVIKAP 1067
 DB 1393 REATTEVGPQIGVKSESSMEPKAAEEAEAGSEQKRDKD--AGTDKNPPTAPAVEVVEKPP 1450
 QY 1068 HAPDPSAFSYPAPPGHPLPLGLHDTARPLPRPPTISNPPLISSAKHPSVLEROIQAISQ 1127
 DB 1451 -APEKNKS-----KRGSRNSRLAVDSASLKN-----VDAAVSPRGAAAQAGERES 1497
 QY 1128 GMSVOLHVPYSEHAKAPVGTWGLPLPMDP-----KKLAPFSG-----VKQEQ-- 1172
 DB 1498 GV-VAVSPEKSESPOKEDGLSOLKSDPVDPKPEKEDVSASGSPSEATQAKQWLELQ 1556
 QY 1173 -----SPRQAGPPESLGVPVTAQEAASVLRGTALGSV----- 1203
 DB 1557 AVEHLAKLAESAASAAKYADAPGELA--PEDRKPAHQASETELAAGIISIIINDISGEPE 1614
 QY 1204 -----PGGSIKGIPISTRV-----PSDSAITYRGSITHGTPADVLYKGTI--TRIIGD 1250
 DB 1615 NFPAPPPYGESQTDLOPPAGAQAALQPS-----EGMETDEAVSGILLETAAATES 1664
 QY 1251 S-----PSRLDRGREDSLPKGHVIVYEGKKGHVLSYEGGMSVTQCSKEDGRSSS 1298
 DB 1665 SRPPVNPAPPSAGPTDTKEARGNSSETSHSPEAKGSK-----EVEVTLVRKDKGRQ-- 1716
 QY 1299 GPHETAAPKRTYDMMGEVRGIRAISSASIEGLMGRAIPPERHSPHLLKQHHIRG-----SI 1355
 DB 1717 -----KITRSRRKRNNTKKV-----VAPVESHP-----ESNOAQGESPA 1752
 QY 1356 TQIGIPRSVYEAQEDYLRRKELLKREGTPPPPPPPPSRDLTEAYKTOALGPLKPAHEGLV 1415
 DB 1753 NEGTTVQHPAPQ-----EKOSEKPHSTPPQOSTSDLSKIPSTP----- 1792

```
QY 1416 ATVKEAGRSIHETPRELHRT-----PELPLAPRLKESGSIOTGTLKVDYTGASTTGSKK 1470
Db 1793 -----NSQEIISVEBETPTKASVPPDLPPPPQ-----AFVDEBPQA-----R 1830
QY 1471 HDVRSLSIGSPGRFP--PVHPLDVMADARALERACVEESLSKSRPGTASSSSGGSIARGAPV 1528
Db 1831 FRVHSIESDPVTPPSDPSIPIPTLSV-----TAAKLSPVASG--- 1870
QY 1529 IVPGLKPRQSP-----LTYBDHGAPFAGHLPRGSPVTMRPTPRLOEGSLSS---SK 1578
Db 1871 -----GPHQSPPTKVTEMITRQE-----EPRAQSTPSPALPDPTKASDVDTIS 1913
QY 1579 ASQDRKLTSTPREIAKS-----PHSTVPEHHHPHPISPYEHLLRGVSGVDLVR 1625
Db 1914 SSTLRKILMDPKVYSATSVTSTVTATIAEPVSAAPCLHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDPTSI PRGIPLDAAAAYLPRHLAPNTYPLHYPPYLIRGYPDPTAALENRQIT 1685
Db 1968 KTAAPPVTNNEISOASEVLVNADEKVPVIAPIKIT-----SVISRMFVSDIDLENSQKI 2020
QY 1686 INDYITSQQMHNTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745
Db 2021 -----TLAKPAPQTLTGL-----VSALTGLVNVSLVP-VNALKG 2053
QY 1746 PTPGTATAMDRLAYLPTAPQPFSSRHSSSPLSPGGPTHLTKTPTTSSSERERDRDRD 1805
Db 2054 PVKGSVTTLKSLVS-----TPAGPVNVLKGPV----- 2080
QY 1806 RDBEREKSILTSTTVEHAPI-----WRPTEQSSGS-----SGSS 1841
Db 2081 -----NVLTGPNVLTTPVNAATGTVNAAPGTVNAASAVNATASAVTVTAGAVTAAS 2133
QY 1842 GGGGGS-----SSRPASHSHAHQHPISPRTODALQOPRPSVLHNTGMKGIIIT 1888
Db 2134 GGTATGTGTVMAGAVIATSKCKQASANENSFRHPGMPVIDDRPA---DAG-SGAGL 2189
QY 1889 AVEPSKPTVLRSPS---TSGPVRPAA---TFPPATHCPLGGTLGGVYPTLMEPVLVPKGA 1942
Db 2190 RVNTSGVLLSYGQKTEGPQISAKISQIPPAS-----AMDIEFQGSVSKSQVKPDS 2243
QY 1943 PRVARP--ERPRADTGHAFLA-----KPPARSGLEPASPSPKSGEPREL--- 1984
Db 2244 VTASQPPSGPQAPAGYANVATHSTLVLTAQTVNASFVITSSVK-ADRPSSL-EKPEPIHLS 2301
QY 1985 -----VPPVSGHATARTPA----- 1999
Db 2302 VSTPVTQGGTVKVLTOGINTPPVLVNVQLVLTFSIVTNNKCLADPTLKIETKVLQPANL 2361
QY 2000 -KNLAPHASPDPPAPASADP--HREKTQSKFP--SIQELRLSLGYHGSSYSPEGYE 2054
Db 2362 GSTLTTPHH---PPALPSKLPTEVNVHVPSPGSI PADRTVSHLAAAKLDAHSRPSGPGPS 2417
QY 2055 PVSFVSPSLSLTHDKGL-----PKH--LEBLDKSHL-- 2082
Db 2418 SFPRASHSPSTASTALSTNATVMAAGIPVPOQFISSIHPEQSVIMPPHSITQTVLSHLS 2477
QY 2083 EGELRPKQPC---PVKLGCEAAHLPHLRPLPESQSSSPLLQTPGVKGHORVVTLAQH 2138
Db 2478 QGEVRNMTPLPSITVSIIRPEALHSR-APL---QP-----QIEVRA-- 2516
QY 2139 ISEVITQDTRHRHPQOLSAPLPALYLSFFGASCFLVLDLRPPSD---LXLPPPDHGAPAR 2195
Db 2517 -----PQASITQPAP-----AGVPALASQHPPEEVHYHLPVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGGEGDIEPVSPPEGM 2231
Db 2558 SEVLVMQSEYRLHPYTPVRDVRIMVHPHTVAVSEQPR-----AADGVVKKVPVASKA 2608
QY 2232 TEPGHRSAYVPLYLLYRDGEQTEPSRMGSKSPGNTSOP 2268
Db 2609 PQ-----QPGKEAAKTPDAAKAP 2626
```

```
RESULT 29
AAM66665
ID AAM66665 standard; protein; 2665 AA.
XX
AC AAM66665;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26971.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US0000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PI WPI; 2001-488900/53.
XX
DR
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX
PS Example 4; SEQ ID NO 26971; 659pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 2665 AA;
```

```
Query Match 3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

QY 71 EFQPGNRSQELHLRPE-----SHSYLPELQKS-----EMEFIESKR---PRIELLP 114
Db 525 ERKSGQEKSHSVNTEBKIGIDIDHTQSYRKQMEQSRKQKQMEMEIAKSKFGSPKDV-- 582
QY 115 DPLLRSPLATQCPAGSEDLTKDRSLTGKLRPVSPPSPHTDPELELVPRLSKEELIQ 174
Db 583 DEYERSLVHEVGKP--PDVTD-----SPPSKK-----K 611
QY 175 NMDRVVDREI--TWVEQOISKLKKKQOOLEBEAAKPP-----EPEKPVSPPIPIES--- 221
Db 612 RMDHVDFDICTKERNYRSRQISEDSERTGSGSPSVRHGSHFDEDPIGSPRLLSVKGSP 671
QY 222 -----KHSRLVQIYYDENRK--AAAHRILEGGLGQVPELPLYNQPSDTRQ 265
Db 672 KYDEKVLPSYNTVRESLSKFNPDSSRRQEMADMAKIKLSVLNSELNRM---DSOM 727
QY 266 YHENIKINAMRKLLIYFKRRNHARKQWKQFCQRYDQJLMEALEKKVEIENNPR--- 322
Db 728 KODAGRFVSPFNII---KRDSLKRKSRVDL--EPGEVPSDSDEGSHKSHSPASAL 781
```



```

Db      2517  -----PQRASTQAP-----AGVPALASQHPPEEVVYHLPVARATAPVQ 2555
Qy      2196  GS-----PHSEGGKRSPEPNKTSVLGGGEDGIEPVSPPEGM 2231
Db      2558  SEVLVMQSEYRLHPYTPVPRVIMVHPHTAVTSQPR-----AADGVVKVPPASKA 2608
Qy      2232  TEPGHSRSVYPLLRYDGEQTEPSRMGSKSPGNTSQP 2268
Db      2609  PQ-----QPGKEAAKTPDKAAP 2626

RESULT 30
AAM54270
ID      AAM54270 standard; protein; 2665 AA.
XX
XX      AAM54270;
XX
XX      05-NOV-2001 (first entry)
XX
XX      Human brain expressed single exon probe encoded protein SEQ ID NO: 26375.
XX
XX      Human; brain expressed exon; gene expression analysis; probe; microarray;
XX      Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX      Homo sapiens.
XX
XX      WO200157275-A2.
XX
XX      09-AUG-2001.
XX
XX      30-JAN-2001; 2001WO-US000667.
XX
XX      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
XX      30-JUN-2000; 2000US-00608408.
XX      03-AUG-2000; 2000US-00832366.
XX      21-SEP-2000; 2000US-0234687P.
XX      27-SEP-2000; 2000US-0236359P.
XX      04-OCT-2000; 2000GB-00024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX      WPI; 2001-483446/52.
XX
XX      Single exon nucleic acid probes for analyzing gene expression in human
XX      brains.
XX
XX      Example 4; SEQ ID NO 26375; 650pp + Sequence Listing; English.
XX
XX      The present invention provides a number of single exon nucleic acid
XX      probes which are derived from genomic sequences expressed in the human
XX      brain. They can be used to measure gene expression in brain cell samples,
XX      which may enable the diagnosis and improved treatment of nervous system
XX      diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX      epilepsy and cancers. The present sequence is a protein encoded by one of
XX      the probes of the invention
XX
XX      Sequence 2665 AA;

Query Match      3.9%; Score 518; DB 4; Length 2665;
Best Local Similarity 19.3%; Pred. No. 9.2e-20;
Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;

Qy      71  EFQGNERSQELHLRPE-----SHSYLPGLGKS-----EMEFIESKR---PRLELLP 114
Db      525  ERKSGQEKSHSVTEEKIGIDIDHTQSYRKQMEQSRKQKQMEMEIAKSEKFGSPKDV-- 582

Qy      115  DPLLRPSPPLATQPGAGSEDLTKDRSLTGKLPVSPSPPHTDPELELVPPLRSKEELIQ 174
Db      583  DEYERRSLNVEVGKFP--PQDVTTD-----SPFSKK-----K 611

```

Db 1498 GV-VAVSPKSESPQKEDGLSSOLKSDPVDPOKEPEKEDVSASGSPPEATQLAKOMELBQ 1556
QY 1173 -----SPRGAGPPESELGVPPTAQASVLRGTALGSV----- 1203
Db 1557 AVEHIAKLAESAASAYKADAPGLA--PEDRDKPAHQASETELAAGIINDISGEPE 1614
QY 1204 -----PGSGITKIGIPSTRV-----PSDSAITYRGSITHGPADVLYKGTI-TRIIGED 1250
Db 1615 NFPAPPYPGESQTDLPAGAAQLPSE-----EGMETDAVSGILETEAATES 1664
QY 1251 S-----PSRLDGRDLSLPGKHVIVGKKGHVLSEGGMSVTCQSKEDGRSS 1298
Db 1665 SRPPVNPAPPSAGTDTKZARGNSSETSHVPEAKGSK-----EVEVTLVRKDKGRQ-- 1716
QY 1299 GPPHETAAKRTYDMMEGRVGRRAISSAIEGLMGRALPERRHSPHLKQEHIRG---SI 1355
Db 1717 -----KTRRRKRNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
QY 1356 TOGIPRSYVEAOEDYLRRBAKLLKREGTPPPPPPSRDLTEAVKTOALGPLKLPKPAHEGLV 1415
Db 1753 NEGTTVQHPEAQ-----BEKQSEKHPSTPPQSCSTSDLSKIPSTE----- 1792
QY 1416 ATVKEAGRIHEIPRELKHT-----PELAPRPLKKEGSIQTGTPKLYDTGASTTGSKK 1470
Db 1793 -----NSSQEISVEERTPTKASVPPDLPPPQP-----ADVDEPQA-----R 1830
QY 1471 HDVRSLLIGSPGRTP--PVHPLDMDADARALERACEYESLKRPGCTASSGSGSIARGAPV 1528
Db 1831 FRVHSIIESDPVTPSPDSIPITPLSV-----TAAKLSPPVASG--- 1870
QY 1529 IVPCLKPQSP-----LTVEDHCAPFAGHLPRGSPVTREPTPRLQEGSLSS---SK 1578
Db 1871 -----GIPQSPPTKVTEMITQSE-----EPRAQSTPSALPDPTKASDVDTIS 1913
QY 1579 ASQDKLTSTPREIAKS-----PHSTVEHPHPPTISPYEHLRLRGVSGVDLYR 1625
Db 1914 SSTLRKILMDPKVVSATSVTSTVTAIAEPVSAAPCLHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDTSIPRGIPLOAAAYLPHLAPNPTYPHLYPPLRYGYPDTAALNRQIT 1685
Db 1968 KTAPPVTNNESEIQASEVLVAADKEKVPVIAPIKT-----SVISRMVPSIDLENSQKI 2020
QY 1686 INDYITSQOMHNTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745
Db 2021 -----TLAKPAQTLTGL-----VSALTGLVNVSLVP-VNALKG 2053
QY 1746 PTFGTATAMDRLAYLPTAPQPFSSRHSSPLSPGPGTHLTPTTTSSSERERDRDRD 1805
Db 2054 PVKGSVTTLKSLVS-----TPAGFVNVLKGPV----- 2080
QY 1806 RDREREKSLTSTTTVEHAPI-----WRPCTEQSSGS-----SGSS 1841
Db 2081 -----NVLTGPNVNLTPVNAVTVNAAGPTVNAASAVNATASAVTVTAGAVTAAS 2133
QY 1842 GGGGGS-----SSRPASHSHAHQSPISPRQTQDALQORPSPVLHNTGMKIIT 1888
Db 2134 GGVATTGTVTWAGAVIATSTCKQASANENSRHFGSPMPVIDDRPA---DAG-SGAGL 2189
QY 1889 AVEPSKPTVLRSTS---TSSPVRPAA--TFPPATHCPLGGTLDGVYPTLMEPVLPLPKA 1942
Db 2190 RVNTSEGVLLSYSGQKTEGPORISAKISQIPPAS-----AMDIEFQQSQSVKQKPS 2243
QY 1943 PRVAPR--ERRADTGHFLA-----KPPASGLEPASPSPKSEPRPL--- 1984
Db 2244 VTASQPPSGPQAPAGYANVATHSTLVLTAQTNASPVISSVK-ADRPDL-EXPEPHLS 2301
QY 1985 -----VPPVSGHATARTPA----- 1999
Db 2302 VSTPVTQGGTVKVLTOGINTPPVLVHNLVLPFISIVTTNKKLADPVLTKIETKVLQPANL 2361
QY 2000 -KNLAPHASPPPPAPPASADP--HRETKQSKPF--STQBLELSLGVHGSYSPEGVE 2054

Db 2362 GSTLTPHH-----PPALPSKLPTENVNHPSPGSPADRTVSHLAAAKLDAHSRPPSPGPGPS 2417
QY 2055 PVSVPSSSLTHDKGI-----PKH--LEELDKSHL- 2082
Db 2418 SPPRASHSPSTASTALSTNATVNLAAAGLPVPOFISIIHPEQSVIMPPHSITQTVSLSHLS 2477
QY 2083 EGELRPKQFG-----PVKLGEAAHLPLRLPSPESQFSSPPLLQTAFGVKGHQVVTLAQH 2138
Db 2478 QGEVRMNTPTLSITYSIRPEALHSPR-APL---QP-----QQIEVRA-- 2516
QY 2139 ISEVITQDTRHHPOOLGAPLPAPLYSPFGACPVLDLRPPSD---LYLPPPDHGAPAR 2195
Db 2517 -----PORASTPQAP-----AGVPALASQHPPEEVEVHLPVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGGEGDIEPVSPPEGM 2231
Db 2558 SEVLVWQSEYRLUHPVTVPDVRIMVHPHTAVSEQPR-----AADGVVKVPASKA 2608
QY 2232 TEPFGHSRAVYPLLRYDGEQTEPFRSMGSKSPGNTSQP 2268
Db 2609 PQ-----QPGKEAAKTPDAKAAP 2626

RESULT 31
ABG48336
ID ABG48336 standard; peptide; 2665 AA.
XX AC ABG48336;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID NO 26984.
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX PN WO200157273-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488898/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX PS Claim 27; SEQ ID NO 26984; 658pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. ABG47348-ABG59930 represent human

Qy	1842	GGGGG	-----SSRPASHSHAHQHSPTSPRTQDALQORPSVLHNTGKMGIIIT	1888
Db	2134	GGVTAITGTVTMAAGVIAPISTKQKQASANENSRPHGSMPIVDDRPA	---DAG-SGAGL	2189
Qy	1889	AVFSPKPTVLRST	---TSFVPRPA---TFPPATHCLPGTGLDGVYFTLMEPVLLPKEA	1942
Db	2190	RVNTSEGWLLSYSGQTEGPQRISAKISQIPAS	-----AMDIEFOQSYSKQVXPDS	2243
Qy	1943	PRVARP	---ERPRADTGHAFLA-----KPPARSGLEPASSPSKSEPRPL	1984
Db	2244	VTASQPPSKGPQAPAGYANVATHSTVLTAQYNASPIVSVK	ADRESL-EKPEPIHLS	2301
Qy	1985	-----VPPVSGHATARTPA	-----	1999
Db	2302	VSTPVTQGGTVKVL	TGINTPPVLVHNLVLTFSITVTTNKKLADPVLTKIETKVLQPANL	2361
Qy	2000	KNLAPHHASDPDPAPASADP	---HREXTQSKPF---SIOELELSLGYHSSYSPGVE	2054
Db	2362	GSTLTPEHH	-----PPALPSKLPTENVHVPSPGSPADRTVSHLAAKLDHSPSPGSPS	2417
Qy	2055	PVSPVSPSLTHDKG	-----PKH---LEELDCKSHL	2082
Db	2418	SFRASHPSSTASTALSTNATVNLAAIGIVPQFISIIHPSQSVMPHPSITQTVSLSHLS	2477	
Qy	2083	EGELRPKQPC	---PVKLGEAAHLPLRLPESOPSSPLLTQAPGVKGQRVVVTLAQH	2138
Db	2478	QGEVRNTPITLPSITYSIRPEALHSPR	APL---QP-----QQIEVRA	2516
Qy	2139	ISEVITQDYTRHHQQLSAPLAPLYSPFGASCPVLDLRRPSPD	---LYLPPPHDGAPAR	2195
Db	2517	-----PQRASTPQAP	-----AGVPALASQHPPEEVHYHLVVARATAPVQ	2557
Qy	2196	GS	-----PHSEGGKRSPEPNKTSVLGGGEGDTEPVSPPEGM	2231
Db	2558	SEVLVMSQYRLHPYTVPRDVRIMVHPHTAVSEQPR	-----AADGVKVPVPASKA	2608
Qy	2232	TEPGHRSRAVYLLRYDRGQTEPSRMGSKSPGNTSQP	2268	
Db	2609	PQ	-----QFGKEAATPDAKAAP	2626
RESULT 32				
AA002259				
ID	AA002259	standard; protein; 2665 AA.		
XX				
AC	AA002259;			
XX				
DT	09-OCT-2001	(first entry)		
XX				
DE	Peptide #941	encoded by probe for measuring human breast gene expression.		
XX				
KW	Probe; human; breast disease; breast cancer; development disorder;			
XX	inflammatory disease; proliferative breast disease; non-carcinoma tumour.			
OS	Homo sapiens.			
XX				
PN	W0200157270-A2.			
XX				
PD	09-AUG-2001.			
XX				
PF	29-JAN-2001; 2001WO-US000661.			
XX				
PR	04-FEB-2000; 2000US-0180312P.			
PR	26-MAY-2000; 2000US-0207456P.			
PR	30-JUN-2000; 2000US-00608408.			
PR	03-AUG-2000; 2000US-00632366.			
PR	21-SEP-2000; 2000US-0234687P.			
PR	27-SEP-2000; 2000US-0236359P.			
PR	04-OCT-2000; 2000GB-00024263.			
XX				
XX	(MOLE-) MOLECULAR DYNAMICS INC.			
PA				
XX				

PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI; 2001-476286/51.
XX	Novel single exon nucleic acid probe used to measuring gene expression in a human breast.
PT	Claim 27; SEQ ID NO 10999; 322pp; English.
PS	
XX	The present invention relates to novel single exon nucleic acid probes (see AA100010-AA110067). The present sequence is a peptide encoded by one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridises at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer, disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 2665 AA;
	Query Match 3.9%; Score 518; DB 4; Length 2665;
	Best Local Similarity 19.3%; Pred. No. 9.2e-20;
	Matches 505; Conservative 315; Mismatches 863; Indels 934; Gaps 120;
Qy	71 EFQPGNRSQELHLRPE-----SHSYLPGLKGS-----EMEFIESKR---PRLELLP 114
Db	525 ERKSGQKSHSVNTEEKIGIDIDHTQSYRKQMEQSRKQKQMEMEIAKSEFGSPKKDV-- 582
Qy	115 DPLLSPPLATGPAGSEDLTKDRSLTKGLEPVSPSPPHPTDPELVPRLSKKELIQ 174
Db	583 DEYERSLVHEVGKP--PDVTTD-----SPPSKK-----K 611
Qy	175 NMDRVDRRI--TWVEQOISKKKQOOLEEAAKPP-----BPEKVPSPPIES----- 221
Db	612 RMDHVDFTCTKRENYRSSQIISEDERTGSGSVRHGSHFHEDEDIGSPRLLSVKGSP 671
Qy	222 -----KHRSLVQIIYDENRKK--ABEAHRILEGLGPQVELPYNQPSDTRQ 265
Db	672 KVDEKVLPSVNTVREESLKENPYDSSRREQMADMAKILSVLNSEDELNRW---DSQM 727
Qy	266 YHENIKINQMRKKLILYFKERNHARKQKFCORYDQLMEALEKKVVERIENPPR--- 322
Db	728 KODAGRFDVSPFNSII---KRDLSLRKRSVRDL---EPGEVPSDDEGEKSHSPASAL 781
Qy	323 -----RAKESKVREYVEKQFPEIRKORELQERMQSRVQGSGSL-----SMSAAR 367
Db	782 YESSLSFLLRDREDKLERDERLSSSLERNKFYSFALDKTIIPDTYKALLERAKSLSR 841
Qy	368 SEHEVSEIIDGLSE-----QENLEKQMRQLAVIPPMPLYDAQQRKIFINNMGLMADPMKVY 423
Db	842 EEN--WSFLDWDSPANFRNNKKEKVDSPAPRPIPSWYMKKKIR--TDSGKMDDKKEDH 898
Qy	424 KDRQVMNWSQEKETTFREKPMQHPKNGFLIASFLERKTVABCVLYYYLTKKNYKSLV 483
Db	899 KEEE-----QERQELFASREL-HSSIFEQDSKRLQ-----HLRKEEDSDFIS 940
Qy	484 RESYRRRGKSQQQQQQQQQQQQQQQQQ-----MPRSSOEKDEKEKEKEAEKEE 533
Db	941 GRIY---GK---QTSEGANSTTDSIQBPVVLTHFRFMELTRMQOKEKEKQOKEPEVKQE 994
Qy	534 EKPEVNDKEDLLKEKTDTSGEDNDEKEAVASKRKTANSQGRRKGRITRSVANEANSE 593
Db	995 ---DTEN-----HPKTPESAPENK-----SELKTP 1017
Qy	594 EAITPQQSAELASWELNESSRWTEEMETAKGLLEHGRNWSAIARMVGSKTVSQCKNFY 653
Db	1018 PSVGPPSV-----TVVTILESAP-----SALEKTTGDKTV----- 1046

QY 654 FNYKKQNLDEILQOHLKMEKERNARRKKKAPAAASEEAPPPVVEDEMEASGVGN 713
Db 1047 -----EAPLVEETKVBATVSBEAKPAPADA 1075
QY 714 EEMVVEEAEALHAGNEVREGCSGPATVNNSSDTSIPSPHTEAAKOTGONGKPPATL 773
Db 1076 PVQLSQV -----DLPFGA -----DPKEAAMMP -----AGVEGSSGDQPP-YL 1114
QY 774 GADGPPGPPPTPRRTSRAP--LEPTPASEATGAPTPPAPPSPSPAPPVVPVPEKEEET 831
Db 1115 DAK -----PPTPGASFQAESNVDPEDPS -----TOPLSKPAKSEANEPAEKPDAT 1163
QY 832 AAAPP -----VEGEEOKPPAAELAVD-----TKABEPVKSECTEBAEAGPAKGDAAEA 883
Db 1164 ADAEPDANOKAEAPESQPPASDLEVPVAAAKKPNKSKRSTPVQAAAVSIVEKPV 1223
QY 884 EATAEGALKAKEGEGGRATTAKSGAQODSDS -SATCSADEVDDEAGGDKNRLSPRP 942
Db 1224 TRKSERIDREKLKRSNPRGEAQKLELKMEEAKI TRTASKNSAADLE -----HPBP 1275
QY 943 SL-LTPT-----GDPRANASPOK-PLD-----LKLKQRAAAALPIPIQVTK 980
Db 1276 SLPKSTRNRNVSVAWGD-HENRSPVKEPVEQPRVTRKLERELQEAAPV--TTPR 1332
QY 981 VHEPPR-----EDAAPTKPAPPAPP-----PONL-----QPESDAP 1012
Db 1333 RGRPPKTRRRADBEENEAEPAETLKPPEGWRSPRSQKTAAGGGPOGKKGNPKVDAT 1392
QY 1013 QOPGSSPRG-----KRSPPADPADKFAFAEAQKLPDPPCWTSGLPFPVPREVTKASP 1067
Db 1393 RPEATTEVGQIGVKESSEMEPKAAEEAGSEQKDRKD--AGTDKNPPTETAPVEVVEKKP 1450
QY 1068 HAPDPSAFSVAPPGLPLGLDHTARVPLPRPTTISNPPPLISSAKHPSVLEROIGAIQ 1127
Db 1451 -APEKNKS-----KGRSRSNRLAVDKSASLN-----VDAASVPRGAAQAQERES 1497
QY 1128 GMSVQLHVPVSEHAKAPVGVTVGLPLPMDP-----KULAPFSG-----VKQEQOL-- 1172
Db 1498 GV-VAVSPKSESPOKEDGLSSOLKSDPVDPDKEPEKEDVSAGSPSPPEATLAKQOMELEQ 1556
QY 1173 -----SPGQAGPPESLGVPTAQEASVLRGTALGSV----- 1203
Db 1557 AVEHIAKLAEASAAKADAPGLA--PEDRDKPAHQASETELAAGIISIINDISGEPE 1614
QY 1204 -----PGSITKIGIPSTRV-----PSDSAITYRGSITHGTADVLVYKGTI--TRIIGED 1250
Db 1615 NFPAPPYPYGESQTDLOPPAGAALQOSE-----EGMETDCAVSGILETEATES 1664
QY 1251 S-----PSRLDRGREDSLPKGHVIVYEGKKGHVLSEGGMSVTCQSKEDGRSSS 1298
Db 1665 SRPPVNPADPSPAGTDTKEARGNSSETSHSPEAKGSK-----EVEVTLVRKDKGRQ-- 1716
QY 1299 GPPHETAAKRTYDMMEGRVGRASISIEGLMGRALPPEHRSPhLKEQHIRG---SI 1355
Db 1717 -----KTRSRKRNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1752
QY 1356 TOGIPRSYVYEAQDYLRRBAKLKRGCTPPPPPPSRDLTEAYKTOALGPLKPKAHEGLV 1415
Db 1753 NEGTTVQHPEAQ-----BEQSEKPHSTPPQSCSDLSKIPSTE----- 1792
QY 1416 ATVKEAGRSIHEIPRELHRT-----PELPLAPRLKEGSIITOGTPLKYDTGASTTGSKK 1470
Db 1793 -----NSSQEISVEERTPTKASVPPDLPPPPOP-----APVDEEPOA-----R 1830
QY 1471 HDVRSLIGSGRFTFP--PVHPLDVMADARALERACEESLKRPGTASSGSGSIARGAPV 1528
Db 1831 FRVHSIIESDPVTPPSDPSIPTPLPSV-----TAAXLSPPVASG-- 1870
QY 1529 IVPCLKPQSP-----LTVEDHGAFFAGHLPRGSPVTMRPTPLQEGSLSS---SK 1578
Db 1871 -----GIPHQSPPTKTEWITROE-----EPRAQSTPSPALPDDTKASDVNTS 1913
QY 1579 ASQDRKLTSTPREIAKS-----PHSTVPEHHHPHPISPYEHLLRGVSGVDLYR 1625

RESULT 33

ABG36319

ID ABG36319 standard; peptide; 2665 AA.

XX AC ABG36319;

XX XX

DT 19-AUG-2002 (first entry)

XX

DE Human peptide encoded by genome-derived single exon probe SEQ ID 25984.

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

Db 1914 SSTLRKILMDPKYVSATSVTSTVTTAIAEPVSAAPCLHEAPPPVD-----SKPLEE 1967
QY 1626 SHIPLAFDPTSIPRGIPLDAAAAAYLPHRLADNPPTYPLHYPPVLYIRGYDPDTALENRQTI 1685
Db 1968 KTAPPVTNNSEIQASEVLVAADKEKVPVIAPIKTI-----SVISRMVPSIDLENSOKI 2020
QY 1686 INDYITSOQMHNTATAMAOADMLRGLSPRESSLALNAAAGPRGIDLSQVPHLPVLPV 1745
Db 2021 -----TLAKPAQTLTGL-----VSALTGLVNVSLVP-VNALKG 2053
QY 1746 PTFGPTATMDRLAYLPTAPQPFSSRHSSSPDLSPGQPTHLTRPTTSSSERERDRDRD 1805
Db 2054 PVKGSVTTLKLSVS-----TPAGFVNVLKGPV----- 2080
QY 1806 RDREREKSLTSTTTTVEHAPI-----WRQTEOSSGS-----SGSS 1841
Db 2081 -----NVLTPGVNVLTTTPVNAATVGTVNAAPVNAASAVNATASAVTVTAGAVTAA 2133
QY 1842 GGGGGS-----SSRPASHSHAHQSPISPRQDALQORPSPSVLHNTGMKIIT 1888
Db 2134 GGVATTGTVTWAGAVIAPSTCKQKQASANENSRPHGSMPIVDDRPA---DAG-SGAGL 2189
QY 1889 AVEPSKPTVLRSTS---TSSPVRPAA---TFPPATHCPLGGTLDGVYPTLMEVLLPKEA 1942
Db 2190 RVNTSEGVLLSVSGOKTEGPORISAKISQIPPAS-----AMDIEFQOSVSKSQVKPD\$ 2243
QY 1943 PRVARP--ERRPADTCHAEFLA-----KPPARSGLPEPASSPSKGSPPRL--- 1984
Db 2244 VTASOPPSKGPQAPAGYANVATHSTLVTAQTVNASPVISSVK-ADRP\$U-EKPEPIHLS 2301
QY 1985 -----VPPVSGHATITAP----- 1999
Db 2302 VSTPVTQGTQVKVLTOGINTPPVLVHNLVLTIPSIVTTNKKLADPVTLKIETKVLPANL 2361
QY 2000 -KNLAPHASPPPPAPPASADP--HREKTQSKPF--STQELERLSLGVHSGSYSEGEVE 2054
Db 2362 GSTLTTPHH---PPALPSKLPTENVHVPSPGSI PADRTVSHLAAAKLDAHSPPSPGPGPS 2417
QY 2055 PVSVPSSPSLTHDKGI-----PKH--LEELDKSHL- 2082
Db 2418 SPFRASHPSSTASTALSTNATVNLAAIGIPVPOFISIHPEQSVIMPPHSITQTVSLSHLS 2477
QY 2083 EGELRPKQPG-----PVKLGCEAAHPLRPLPESQSPSSPLLQTAAPGVKHQHVVTLAQH 2138
Db 2478 QGEVRMNTTLP\$ITVSIRPEALHSPR-APL---QP-----QQIEVRA-- 2516
QY 2139 ISEVITQDYTRHHPOOLSAPLPAPLYSPFGASCPLDLRRPSPD---LVLPPPDHGAPAR 2195
Db 2517 -----PORASTPQAP-----AGVPALASQHPPEEVHVLHLPVARATAPVQ 2557
QY 2196 GS-----PHSEGGKRSPEPNKTSVLGGEGDGIIEPVSPPEGM 2231
Db 2558 SEVLVWQSYRLHYPTVPRDVRIMVHPHVTAVSEQPR-----AADGVKVPVPA\$KA 2608
QY 2232 TEPGHRSR\$AVYLLYRDGQTEPSRMGSKSPGNTSQP 2268
Db 2609 PQ-----QPGKEAAKTPDAKAAP 2626

tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
 pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 primary ciliary dyskinesia; pulmonary hypertension;
 hyaline membrane disease.

Homo sapiens.
 WO200186003-A2.
 15-NOV-2001.
 30-JAN-2001; 2001WO-US000665.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.
 21-SEP-2000; 2000US-0234687P.
 27-SEP-2000; 2000US-0236359P.
 04-OCT-2000; 2000GB-00024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2002-114183/15.
 Spatially-addressable set of single exon nucleic acid probes, used to
 measure gene expression in human lung samples.
 Claim 27; SEQ ID NO 25984; 634pp; English.

The invention relates to a spatially-addressable set of single exon
 nucleic acid probes for measuring gene expression in a sample derived
 from human lung comprising single exon nucleic acid probes having one of
 12614 nucleic acid sequences mentioned in the specification, or their
 complements or the 12387 open reading frames derived from the 12614
 probes. Also included are a microarray comprising the novel set of probes
 in the novel set of probes which hybridise at high stringency to a nucleic
 acid expressed in the human lung; measuring gene expression in a sample
 derived from human lung, comprising (a) contacting the array with a
 collection of detectably labeled nucleic acids derived from human lung
 mRNA, and (b) measuring the label detectably bound to each probe of the
 array; identifying exons in a eukaryotic genome, comprising (a)
 algorithmically predicting at least one exon from genomic sequences of
 the eukaryote; and (b) detecting specific hybridisation of detectably
 labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 having a fragment identical to the predicted exon, the probe is included
 in the above mentioned microarray; assigning exons to a single gene,
 comprising (a) identifying exons from genomic sequence by the method
 above and (b) measuring the expression of each of the exons in several
 tissues and/or cell types using hybridisation to a single exon
 microarrays having a probe with the exon, where a common pattern of
 expression of the exons in the tissues and/or cell types indicates that
 the exons should be assigned to a single gene; a peptide comprising one
 of 12011 sequences, mentioned in the specification, or encoded by the
 probes/open reading frames (ORF). The probes are used for gene expression
 analysis, and for identifying exons in a gene, particularly using human
 lung derived mRNA and for the study of lung diseases such as asthma, lung
 cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary
 histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 Kargener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 present sequence is a peptide/protein encoded by a single exon probe of
 the invention. Note: The sequence data for this patent did not form part
 of the printed specification, but was obtained in electronic format
 directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match	3.9%;	Score 518;	DB 5;	Length 2665;
Best Local Similarity	19.3%;	Pred. No. 9.2e-20;		
Matches 505;	Conservative 315;	Mismatches 863;	Indels 934;	Gaps 120;
71 EFQGNERSQELHLRPE	-----SHSYLPELGKS-----	EMEFIESKR	-----PRLELPP	114
525 ERKSQEKSHSVNTEKIGIDIDHTQSYRKQWQSRKQWQWEMETAKSEKFGSPKQDV--	-----		-----	582
115 DPLLRPSLLATGQAGSEDLTKDRSLTGKLEPVPSPSPHPTDPELELVPPLSLKEELIQ	-----		-----	174
583 DEYERRSLVHEVGK--PDVTDD-----	-----SPPSKK-----		-----	611
175 NMDRVREI--TMVEQISKLKKQOOLEEAAKPP-----	-----EPEKPVSPPIES-----		-----	221
612 RMDHVDFTCTKRNRYSSRQISSEDSERTGSGSVRHGSHFDEDDPGISPLLVSVKSP	-----		-----	671
222 -----	-----KHSILVQIYDENRKK-----	AEAAHRIEGLGQVLPVLPYQPSDTRQ	-----	265
672 KYDEKVLVPSNITVREESLKNFYPDSSRREQWADMAKILSVLNSEDELNRW----	-----DSQM	727		
266 YHENIKINQAMRKKLILYFKRNHARKQWKQFCORYDQLEALKEKKVERIENRPR----	-----	322		
728 QODAGRFDVSPFNSII-----	-----KDSLRKRSVRDL-----	EPGEVPSDSDEDEGSHSHSPRASAL	781	
323 -----	-----RAKESKVREYVEKQFPEIRKORELQERMOSRVGQSGGL-----	SMSAAR	367	
782 YESSRLSFLLRDRDLKRLRDERLSSSLERNKFYSFALDKITTPDTKALLERAKSLSSR	-----	841		
368 SEHVESEIIDLSE-----	-----QENLEKQMRQLAVIPPLYDADQOQRIKFINMNGLMADPMKY	423		
842 EEN--WSFLWDMSRFANFNKKEKVDSPAPRIPSWYMKKKIR--TDSEGMDDKKEDH	-----	898		
424 KDRQVNNWSEQEKETPFREKPMOHPKNGLTLASFLERKTVAECVLYLYLTKNVYKSLV	-----	483		
899 KEEE-----	-----QERQELFASRFL--HSSIIEQDSKRLQ-----	HLERKEEDSDFTS	940	
484 RSYRRRGKSGSQOQQOQQOQQOQQOQQOQ-----	-----MPRSSQEEKEKEKEKEKEKE	533		
941 GRIV---GK---QTSEGANSTTDSIQEPVLFHSRPMELTMOQKEKEKDKPKEVEKE	-----	994		
534 EKPEVENKEDLLKKTDTDTGSDENDEKAVASKGRKTANSQGRKRGKITRSMANEASE	-----	593		
995 ---DTEN-----	-----HPKTPESAPEND-----	SELKTP	1017	
594 EAITPQQAELASLMELNERRWTEBEMETAKGLLEHGNWSAIARMVSGSKTVSOCKNFP	-----	653		
1018 PSVGPPSV-----	-----TVVTLSSAP-----	SALEKTTGDKTV	1046	
654 FNYKKRONLDELLOQHLKMEKERNARRKKKAPAAASEEAAFPVVEDEMEASGVSGN	-----	713		
1047 -----	-----EAPLVEKTVPATVSEAKPASEPAPA	1075		
714 EEMVVEEABALHASNEVPRGCSGPATVNNSSDTESTPSHTEAAKDTGONGPKPPATL	-----	773		
1076 PVEQLEQV-----	-----DLPPGA-----	DPDKAAMMP	1114	
774 GADGPPPGPPTPPRTSRAP--IEPTASEATGATPPAPPSAPPVPPVPEKEESET	-----	831		
1115 DAK-----	-----PPTPGASFQAEENVDPEPS-----	TQLSKPAQKSEANEPAKSPDPT	1163	
832 AAAPP-----	-----VEEGEQKPPAAEELAVD-----	TGKAAEFPVSECTBEAEGPAKDAFAA	883	
1164 ADAEPDANKAEAPESQPPASEDLVDPVPAADKKFNKSKRSTPVQAAAVSIVEKPV	-----	1223		
884 EATAGALKAEKKEGSGRATTAKSSGAPQSDS--SATCSADEVDEAEGDKNRLSPRP	-----	942		
1224 TRKSRIDREKRLKRSNPRGEAQKLELMEAEKTRTASKNNSAADLE-----	-----HPEP	1275		
943 SL-LPPT-----	-----GPRANASPOK--PLD-----	LKOLKORAAAIPIQVTK	980	
1276 SLPLSRTRRRNRVSYYATMGD--HENRSPVKEPVEQPRVTRKLELRELQEAANVP--	-----	1332		

Qy	981	VHEPPR-----EDAAPTKPAPPPP : :	-----PNL-----QPESDAP 1012
Dd	1333	RGRPPKTRRRADBEENEAKPAETLLKPPBGRWSPRSOKTAAGGPGCKKGKNPKVDAT 1392 : :	
Qy	1013	QQGSSPRG-----KSRSAPPADKEAFAEAQKLPGDPFCWTSGLPFPVPPREVIKASP 1067 : :	
Dd	1393	RPEATTEVGPOIGVKESSMPEFKAAEEBAGEQKRDRKD---AGTDKNPFPTAPVEVVVEKKP 1450 : :	
Qy	1068	HAPDPAFSAVPCHPLPLGLHDTARVLPRPPTISNPPLIISAXHPSVLERIQIGAISO 1127 : :	
Dd	1451	-APEKNSK-----KRGSRNSRLAVDKSASLKN----VDAAVSPRGAQAAGERES 1497 : :	
Qy	1128	GMSVOLHVPIYSEHAAPGVPTWGLPLPMPD-----KKLAPFG-----VKQBOL-- 1172 : :	
Dd	1498	GV-VAVSPEKSESPOKEDGLSSQLSDPDVDPDEKEPEKEDVASGSPGEATQALAKAMELEQ 1556 : :	
Qy	1173	-----SPRGOAGPPESLGVPТАQEASVLRTGTALGSV----- 1203 : :	
Dd	1557	AVEHIAKLABASASAAYKADAPEGLA---PEDROKPAHQASETAELAAIGAISIINDIGEPE 1614 : :	
Qy	1204	-----PGCSITKGIPSTRV-----PSDSAITYRGSITHGTPADLVLYKGTI-TRIIGED 1250 : :	
Dd	1615	NFPAPPYPGESQTDLOPPAGAQAOPSE-----EGMETDEAVSGILETEAATES 1664 : :	
Qy	1251	S-----PSRLDRGRESOLPKGHVIVYGKKGHVLVSYGGGWSVTQCSKEDGRSSS 1298 : :	
Dd	1665	SRPPVNAPDPSAGPTDTKEARGNSSETSHSVPEAKGSK-----EVEVTLVRKDKGRQ-- 1716 : :	
Qy	1299	GPHETAAPKRTYDMMEGRVGRAISSAS IGLMGRAIPPERHSPHHLKHOHHIRG---SI 1355 : :	
Dd	1717	-----KTTSRRKRNTNKKV-----VAPVESHYV-----ESNQAGSSPAA 1752 : :	
Qy	1356	TGIPRSYVAOBDYLRRREALKLRCTGTPPPPPSRDLTAYKQTQALGPLKLKAHEGLV 1415 : :	
Dd	1753	NEGTTVOHPEAQ-----EEKOSEKPHSTPQOSTSDLSKI PSTE----- 1792 : :	
Qy	1416	ATVKEAGRSIHETPREBLRHUT-----PELPLAPRLKBSGITOGTPLKYDTGASTTGSKK 1470 : :	
Dd	1793	-----NSSQETISVEERTTKASVPDLPPPPQP-----APDESPOA-----R 1830 : :	
Qy	1471	HDVRSLIGSPGRTFP--PVHPLDVMDARALARACYEESLKRPGTASSGGSGIARGAPV 1528 : :	
Dd	1831	FRVHSIIESPVTPPSDPSIPIPTLSV-----TAAKLSPPVASG--- 1870 : :	
Qy	1529	IYPELGKPROSP-----LTYESDHGAPFAGHLPRGSPVTVMRPFTPRLOEGLSS---SK 1578 : :	
Dd	1871	-----GIPHQSPTKVTEMITROE-----EPRAQSTFSPALPPDTKASDVODTS 1913 : :	
Qy	1579	ASQDRKLTSTPREIAKS-----PHSTVPEHHHPHTSYEHLLRGLGVSGVDLYR 1625 : :	
Dd	1914	SSTLRKILMDPKVYSATSVTSTSVTTAIBFVSAAPCLHEAPPPVD-----SKXPLEE 1967 : :	
Qy	1626	SHIPLAFDPTSIPRGIPLDAAAAAYLPHULAPNPTYPHLYPYVLIIRGYPDATALENRQTI 1685 : :	
Dd	1968	KTAPPVNNSEIQAASEVLVAADKEKAPVIAPKIT-----SVISRMPPVSIIDENSOKI 2020 : :	
Qy	1686	INDYITSQQMHNTATAMAOADMRLGSLPRESSLALNYAAGRPIIDLISQVPHLPVLVP 1745 : :	
Dd	2021	-----TLAKPAPQTLTGL-----VSALTGLVNVSLVP-VNALKG 2053 : :	
Qy	1746	PTPCTPATMDRLAYLPTAPQPPSSRHSSPSLPGGPTHLTTKTTTTSSSERDRDRERD 1805 : :	
Dd	2054	PVKGSVTLTKSLVS-----TPAGPVNVKLGPV----- 2080 : :	
Qy	1806	RDREKERKSIILTSTTTVEHAPI-----WRPGTESQSSGS-----SGSS 1841 : :	
Dd	2081	-----NVLTGPVNVLTTPVNATVGTVNAAPGTVNAASAASNATASAVTVTAGAVTAAS 2133 : :	
Qy	1842	GCGGGS-----SSRPASHSHAHOHQSPISPRTODALQQRPVSLHNTMKGIIT 1888 : :	
Dd	2134	GGVATTGTTVMAGAVIAPTCKKQRASANENSRFHGSMFVIDDRPA---DAG-SGAGL 2189 : :	

Qy	1889	AVESKPTVLRST---	TSSPVRPA-----TFPPATHCPLGGTLDGVYPTLMPEVLVPKEA	1948
Db	2190	RVTSEGWLLSYSGQTEGPQIRISAKISQIPPAS-----AMDIEFQQSVKSQVKPDS	2243	
Qy	1943	PRVARP--ERPRADTGHAFLA-----KPPARSGLPEASSPSKSGSPRPL---	1984	
Db	2244	VTASQPPSKGPQAPAGYANVATHTSLVLTQATYNASPVISVK-ADRPSL-EKPEPIHLS	2301	
Qy	1985	-----VPPVSGHATIAITPA-----	1999	
Db	2302	VSTPVTGGTVKVLTOGINTPPVLVHNQLVLTSTIVTTNKKLADPVTLKTIETKVLPANL	2361	
Qy	2000	-KNLAPHASDPDPAPPASADP--HREKTQSKPF--SIOELELRSLGYHGGSSYSPEGVE	2054	
Db	2362	GSTLTTPH-----PPALPSKLPTEVNHVPSGSPADRTVSHLAAAKLDAHSPRPSGPGPS	2417	
Qy	2055	PVSPVSPSLTHDKGL-----PKH--LEELDKSHL--	2082	
Db	2418	SFPRASHPSSASTALSTNATVMLAAGIPVQFISSIHPEQSVIMPPHSITQTVSUSHL	2477	
Qy	2083	EGELRPKQPG-----PVKLGCEAAHLPHLRPLPESQPSSSPLLQTAPGVKGHQRVVTIAQH	2138	
Db	2478	QGEVMTPTPLPSITYSIRPEALHSFR-APL---QP-----QQIEVPA--	2516	
Qy	2139	ISEVITQDTRHHFQQQLSAFLPAPLYSFPFGASCPLDLRPPSD---LYLPPDPDHGAPAR	2195	
Db	2517	-----PQRASTPQAP-----AGVPALASQHPPEEVHYHLFVARATAPVQ	2557	
Qy	2196	GS-----PHSEGGKRSPEPNKTSVLGGGEDGIEPVSPPEGM	2231	
Db	2558	SEVLVMQSEYRLHPYTPVRDVRIMVHPHTAVSEQPR-----AADGVKVPVPASKA	2608	
Qy	2232	TEPGHRSYAVYLLYRDGEQTEPSRMGSKSPGNTSQP	2268	
Db	2609	PQ-----QPGKEAAKTPDAKAAP	2626	
RESULT	34			
ABB63299				
ID	ABB63299	standard; protein; 2703 AA.		
AC	ABB63299;			
CC	ABB63299;			
DT	26-MAR-2002	(first entry)		
DE	Drosophila melanogaster	polypeptide SEQ ID NO 16689.		
XX	Drosophila;	developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.			
XX	Drosophila melanogaster.			
OS	Drosophila melanogaster.			
XX	WO200171042-A2.			
FN	27-SEP-2001.			
XX	23-MAR-2001;	2001WO-US009231.		
PF	23-MAR-2000;	2000US-0191637P.		
XX	11-JUL-2000;	2000US-00614150.		
PR	(PEKE)	PE CORP NY.		
PA	Venter JC, Adams M, Li PWD, Myers EW;			
PI	WPI; 2001-656860/75.			
XX	N-PSDB; ABL07402.			
DR	New isolated nucleic acid detection reagent for detecting 1000 or more			
XX	genes from Drosophila and for elucidating cell signaling and cell-cell			
PT	interactions.			
PT	Disclosure: SEQ ID NO 16689; 21bp + Sequence Listing; English.			
PS				

XX	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
CC	sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-		
CC	ABBS72072). The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX	Sequence 2703 AA;		
XX	Query Match		
XX	Best Local Similarity 20.1%; Pred. No. 1.9e-19;		
XX	Matches 435; Conservative 187; Mismatches 746; Indels 797; Gaps 103;		
QY	670 KLMEKERNARRKKKAPAA-----ASBEAAPPVVEDEMEASGVSGNEHEMVEEAEA 723		
DB	4 KIKSPOTQOQOGGAPAPAAATPSAGAAPGAATPPT----- 39		
QY	724 LHASGNEVREGSGPATVNNSSDTSIPS-----PHEAAKDTGQNGKPPATLG 774		
DB	40 -----SGPPTPNNSNNGSDPSIQOQOQNAVAPHYGA-----PPPPGS-- 77		
QY	775 ADGPPPPGP-----PTPRRTSR-----APTEPTASEATGAP----- 806		
DB	78 GPGGPGPAPAAVMYHLLHQOQOQHPHPHQQOQHHGHPAPPPGGAPHAHPGVKEY 137		
QY	807 --TPPPAP-----PSPAPPVVPVKEKEEETAAPVVE-----GE 841		
DB	138 THLPHPHPHAYGRYHADPNMDPYRGQLPGGKBP-----QQQHPHQQPQQPQPGPGS 193		
QY	842 EQKPPAAELAVDTGKAEPVSEKTEBAEEP-----AKGDABAAEATAGALKAEK 895		
DB	194 PNRPOQRYIPGPPGQPTFTLSLQSSNPPPPHQRHYANTYDPOQAAASAAAAAQO 253		
QY	896 KEGSGRATTAKSSGAPQSDSATCSADEVDEAEGGDKN-----RLLSPR----- 941		
DB	254 QQAG-GPPPPHGGPPPPHQS-----PYGQOQGWAPPFRPYSPQLGPSQOYR 301		
QY	942 --PSLLTPTGD---PRAN-----ASPQPLDLKOLKORAAI-----PPIQVT 979		
DB	302 TPTPTNTSRGQSPYPPAHGQNSGYSPPSQOQOQOQOQOAGQPGVPFGPPPGTG 361		
QY	980 KVHEPREDAATKPAPAP---PPQNL-----OPESAPQOQSSPRG 1021		
DB	362 Q--QPQQTPTPTQSVSPYQRYPTPPGLPAGGSNHRATYSTHQYPEPNRPMWGSSSP 419		
QY	1022 KSRSPAPPADKEAFAEAOQLPGDPPCWTSGLPFPVPPREVIKASP-HAPDPSAF-SYAP 1079		
DB	420 GSGHPLPASP-HVPPLOQOQPPPPHVSAGP---PP-----SSSPGHAPSPPQPSQAS 471		
QY	1080 PGHPLPL-----GLHD---TARVPLRPPTTISNPPLISSAKHPSVLERQIGAI 1125		
DB	472 PSPHQELIQNSDSSGSAHSGMGGPCTENPQQWMPRTPSPTGSS-----GSR 522		
QY	1126 SQGMSVOLHVPYSEHAK-----APV---CPVTMGLPLMDPKLAPFSGVKQEOISPRQ 1177		
DB	523 SMSPAVAQHPIISRPNASNGSGGPMQOPFVAGGPPMPHPHGMF-GGFPQOQOQOQO 581		
QY	1178 AGPPESLGVPTAQEASVLRGTALGSPVGGISITKGIPTSTRVPSDSATYRGSITHTGTPADV 1237		
DB	582 ANSASSASNSPQOT-----PPAPPPQGMNMMATPPPP---QGAAGGYVMP 629		
QY	1238 LYKGTITRIIGEDSPSLDRGSDSLPKGHVYIEGKKGHLVSYEGMSVTQCSKBDGRSS 1297		
DB	630 HMHGKYK--MGPGQSPGAQGYPPQPOQ-----QYPPGNYPRPQYPPGAYA 674		
QY	1298 SGPPHETAAPKTYDMWEGRVGRTASSASIEGLMGAIPPE---RHSPHIL---KQHHIR 1352		
DB	675 TGPP-----PPPTSQAGAGANSMPGSAQAGGYPGKGMFNHTGQVPPYQWVPPSPQQTVP 729		
QY	1353 GSITQIIPRSYVEAQEDYLRREAKLLKEGTTPPPPPPPSRDLTEAYKTQALGPLKLPKPAHE 1412		
DB	730 G-----GAPCGAMVGNH-----VQKGTPPPPPVVG-----GP-----PPQ 760		
QY	1413 GLVATVKGAGRSIHPIPREELAHTELPPLAPRLKESITQGTPLKYDTGASTTSKKHD 1472		
DB	761 G-----SGSPRLNLYL-KQHLQHKGGYGSPTP-----PQG-PQGYNG--PTGM----- 801		
QY	1473 VPSLIGSCRTPPPVH-----PLDMADARALERACYEESLKSPPGTASSSG 1519		
DB	802 -----HPGMPMGPPHHMGPPHGTNMGPTTTPPOSQMLQ-----GGQPGQCGASG 847		
QY	1520 GSIRGAPVIVPELCKPRQSPLTYEDHGAPFAGHLPRGSPVMTREPTTRELQSG--SLSSS 1577		
DB	848 G-----PESGPEH---ISQDNGISSG--PTGA-AGMHAVTSVVITGPDGTSMD 891		
QY	1578 KASQDKLTSTPRETAKSPHSTVPHHPHIPISPYEHLRLRGVGVDLRYSHIPLAFDPTSI 1637		
DB	892 EVSQOSTLSNASSAGEDPQCTTPKSRKN-----DP----- 922		
QY	1638 PRGIPLDAAAAYLPRHLAPNPTYHLYPPYLRGYPDTAALENQTIINDYITSQOMHH 1697		
DB	923 -----YSOSHLAPPSTSPH----- 936		
QY	1698 NTATAMAQADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPPTPGTPATAMDR 1757		
DB	937 -----PVVMHPPGGGEEYDMSPPNWP-----RPAGSPQV 967		
QY	1758 LAYLTPAQPPSSRHSPLSPGGPTHLTKTPTTSSSR-----ERDRDR-----DRDR 1808		
DB	968 FNVHPVPOEPFRS-----TTTTKKSDSLCKLYEMDDNDPRRRGLDKLR 1011		
QY	1809 ---EREKSLTSTTVEHAPI-----WRPTEQSSGS-----SGSSGGGGSSS 1849		
DB	1012 AFMEERTPTTACPTISKQPLDLYLYIVYKRGFVETKSTWKDLAGLIGASSSA 1071		
QY	1850 RPASHSH-----AHQHSPISPRTQDALQ 1872		
DB	1072 AYTLRKHYTKNLLTFECHRDGIDPLPIIQOVEAGSKKTKATAASVSPGSSNSQDSFP 1131		
QY	1873 QRPVSLVHTGKMG-----IITAVBSKPTVLRSTSTSSPVRAAATPPATHCPL 1921		
DB	1132 APPGSAPNAAIDGYPGPGSPYVAVSGPQDYATAGQMQRPPSQNNPQTTHPGAAAAVA 1191		
QY	1922 GGTLDGVVPTLMEPVLLPKEAPVARPERPRADTGHAFKAPKPARSGLEPASSP----- 1975		
DB	1192 AG--DNI-----SVSNPFEDP-IAAGGPGSGTG-----PGCGGPGFGAASGAGAV 1236		
QY	1976 --SKGSEPRPLVPPVSGHATIAATPA---KNLAPHASPDPPAPPASASDPHREKTQSKPF 2031		
DB	1237 GAVGGGPQHPHPHPHSPHTAAQAAQHQHQHQHQPGLPGFPFPFPQOQOQOQOQPP- 1295		
QY	2032 SIQLELELSLGVHSGSYSP-----GVEPVSPVSSPSLTHDKGLPKHLEELD KSHLEGLR 2087		
DB	1296 -----SVG-GGPPAPQHQGQVFPSP-----QOHVR 1322		
QY	2088 PKQPGPVKLGGEAHLPHLRPLPESQSSSPLLOTAPGVKGHQVRVTLAQHISEVITQDY 2147		
DB	1323 PAAGAPYPPGSG-----YPTVSRTPGSP-YPSQPGAYG-----QY 1358		
QY	2148 TRHHPQQLSAPLAPLYFPFGASCFCVLDLRPPPSDLYL-PP--PDHGAPAGS----- 2197		
DB	1359 GSSDQYNATGPPGQFGQPG-----QYPPQNRNMVPPYPPGEGEAPPTGANQYGPVG 1410		
QY	2198 --PHSEGGKRPPEPKNTSVLGGEGDIEPVSPPEGKTEPHGSRSAIVYLLVRDGEQTEPS 2255		
DB	1411 SRPYSQPPGPGPPTQTVAAGPPAGGAPGAPP-----SSAYP----- 1448		
QY	2256 RMGSKSPGNTSOPPAFFSKLTSNSAMVSKKQEKINKLINTHNRNEPVNISQPCTEIFN 2315		
DB	1449 -TGRFSQDYQPPPPDQSPQRRHDFDK-----SQYPYGNA----- 1486		

Query Match 3.9%; Score 512.5; DB 5; Length 2703;
Best Local Similarity 20.1%; Pred. No. 1.9e-19;

Db 1618 TKALE-----BSLVQEGR--AREQBEKWR-----QDVV-----QEW 1650
QY 610 NESSRWTEEMETAKKLEHGNWGAIAARMVSGKTVSQCKNFYFNKKRQNLDEILOOH 669
Db 1651 QETSPTRPEAGQK-----ELAPW-----EDTSPQDNRYWGRGDEVALEQDTYWR 1698
QY 670 KLMWEKERNARRKKKAPAAASEEAPPPVVEDEEWEASGVSGNEBEMVEEAEALHAS-- 727
Db 1699 ELSCE-----RKWVFPHELDQGGARPHYTERESTFLDEGDDQEVPLRE--HATRS 1749
QY 728 -----GNEVPRCEGSGPATVNNSSDTE-----SIPSPHTEAAKDTGN 765
Db 1750 PWASDFKFOESSFQKLEVERWLAESPVLPEEEDKLTSPFEILISPPASPPWVGQR 1809
QY 766 GKPPATLADGGPPGPPPTPPRTSRAPTEPASEATCAPTP-----PAPPSPSAPP 819
Db 1810 VFSAP---GOESPDPKLMFMKN-----EPTTPSML--ADIPWVPKDRPLPAPLSA 1860
QY 820 PVVPKEEKEETAAAPVVEGEQKPPAAABELAVDTGKABEVPKSECTEAEAGPAK--G 877
Db 1861 PGPTTAPESHTPA--PFGWGTAEYDSV--AAVQEGAA-----ELEGPPYSLG 1906
QY 878 KDAEAAEAATAGALKAEKKEGGSGRA--TTAKSSGAPQSDSDSATSADENVDEAEGDXN 935
Db 1907 KQYKAE-----GEREEGRABADKSSHSKVPKASKHATTEPQTE-----PEQ 1953
QY 936 RLISPRPS-----LLT-----PTGDRANASPOKP--LDLKQL-----966
Db 1954 REPTPYDERSFOYADIEQOMLTGCPACTREPLGAGDWPCCLSTYKEAAGRNTSA 2013
QY 967 -KORAAAIPIQVTKVHEPREDAATKP--APPAPPPQNLQSPESDAPQOQSSSPKGR 1024
Db 2014 EKELSPISGP-KSLQSDTPTFSVAALAGTVPVPRPFGSPMEPSLTPPAVPPRAPI--LSK 2071
QY 1025 SPAPPADKFAFAEA--QKLPDGPCTWSGLPPVPPREVIKASPHADPDPSAFSAPGHP 1083
Db 2072 GSPPLNGILNCSPPRRSPSKESGRSHWDDSTSDSELEKGAQEKEAQSPP--HP 2130
QY 1084 LPLGLHDTARVPLPRPTTNNPPLISSAKHPVLERQIGAISQGMVOLHVPYSEHAKA 1143
Db 2131 IPWG-----SPTL-WPETAHVSPPLDLSHLP-----RPSLDFFASAFGS 2171
QY 1144 PVGPVTMGLPLMDPKKLAP-----FSGVKQEQLSPRGOAGPP-----ESLGVPFTAQSA 1192
Db 2172 SLOPAPPQLPSPAEPSP--APCGSLAFSGDALALAP---GPPTRTRHDEYLEVTKAPSL 2226
QY 1193 -SVLRGTALGSGVGGSIKGIPTSTRVPSDSAITVRSITHTGTPADVLTKGTTIRIGEDS 1251
Db 2227 DSSLPOLPSPSPGAPLLNLPR---PASPULSEGSSSEATTP-----VISSVAERFS 2276
QY 1252 PS-----RLDRGREDSLPKGHVIEYCKGHVLSYEGGMSVTQCKSKEDGRSSGPPH 1302
Db 2277 PSLEAAEQSGELDPGME---PAAHSLWD-----LTPLS-----2307
QY 1303 ETAAPKRTYDMMEGRVGRAISSA-SIEGLMGAIPP-----ERHSPHLKQHHI 1351
Db 2308 --PAPPASLDL-----ALAPAPSLPGMDGILPCHLECSAAATEKSPFQVPSDECA 2358
QY 1352 RGSITOGIPRSYAEQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQALGPLKLKPAH 1411
Db 2359 ANGPTETSP-----NPPGAPAKAENE-----2380
QY 1412 EGLVATVTKAG--RSIHETPREELRHTPELPLAPRPLKSGSITQGTPLKYDTGASTGS 1468
Db 2381 EAAACPAWENGAWPEGAERSSRPDTLLSPQVCPAGGSGPPSSASP-EVEAGPOGCAT 2439
QY 1469 KKHVRLSLGSPGRTFPFVPLDVMADARALERACYEESLKSPGTASSGGSIGARGAPV 1528
Db 2440 EPRPHRGEL-SPSFLNPLPP---SIDRDL--STEEVLVGR-----GRRRRVGGPG 2486
QY 1529 IYPELCKPQSPLYEDHGAFFAGHLPRGSPVTWREPTPLQEGSLSSSKASQDKLTST 1588
Db 2487 TT-----GQPCPVTDETPTTSASDSGSGQSDS-----V 2515

RESULT 38

AAB50362

ID AAB50362 standard; protein; 3118 AA.

XX AC AAB50362;

XX DT 12-MAR-2001 (first entry)

XX DE Human SRCAP.

XX KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;

XX KW CAMP regulatory element; CREB binding protein; CBP; AIFase;

XX KW transcription activation; DEAD box RNA dependent helicase;

XX KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.

XX OS Homo sapiens.

XX PN WO200073467-A1.

XX PD 07-DEC-2000.

XX PF 25-MAY-2000; 2000WO-US014719.

XX PR 27-MAY-1999; 99US-0136620P.

XX PR 25-MAY-2000; 2000US-00579181.

XX PA (UYSL-) UNIV SAINT LOUIS.

XX PI Chrivia J, Yaciuk P;

XX PS WPI; 2001-061545/07.

XX DR N-PSDB; AAC89859.

XX XX Snf2 related CAMP regulatory element (CREB) binding protein (CBP) activator protein, capable of co-activating CREB binding protein, useful for modulating transcription and for affecting viral infection.

XX XX Claim 6; Page 77-86; 103pp; English.

XX CC The present sequence is an Snf2 related CREB (cAMP regulatory element) binding protein (CBP) activator protein (SRCAP) polypeptide. It has Arpase activity and is capable of activating transcription. SRCAP polypeptides are useful for activating transcription in a cell, for enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated activation of transcription in a cell, for treating a patient having a disease involving a function such as insufficient transcription of a gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent helicase, adenoviral DBP protein, beta-actin or a nuclear receptor affected by SRCAP protein. Compounds that modulate SRCAP function, such as antibodies, antisense molecules, polynucleotides or ribozymes, are useful for treating diseases mediated by SRCAP-activated transcription, for example, infection by adenovirus, hepatitis C virus, human immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or hepatitis B virus

XX XX

[illegible]

1676 -TAALENRO-----TIINDYITQ---QMHNNTATAMAQADMURG- 1712
1933 QTLAVLLRQKABGHRVLI FTOMTMDLVLEQLFTHGHLYLRDLGSTRVEQORQALMERF 1992
1713 -----LSPRESSLANYAAGPRGIIDLSQVPHLPVLVPPPTCTPATAMDR----- 1757
1993 NADKRIFCFILSTRSGGVNLTGADTVVPYDSDW-----NPTMDAQADCHRRIG 2043
1758 -----LAYLPTA-----POPFSS 1770
2044 QTRDVHLYRLISERTVEENILKKAQKRMGLDMAIEGNGFTTAYFKQOTIRELFDMPLEE 2103
1771 RHSS-PLSP-----GPHLTKPTTSSSERERDR----- 1800
2104 PSSSSVPSAPEEBEETVASKQTHILEQALCRADEEDIRAATAQAKAEQVAELAEFNENDG 2163
1801 -----DRDRDREREKSIILSTTTVEH----- 1823
2164 FPAGEGEAGRPCAEBEEMSRASQETAALVEQUTPIERYAMKFLASLEBVSREELKQAE 2223
1824 -----APIWR-PGTEQSSSGS--SSGGGGSSSRPASHSHAHQH--SPIS 1864
2224 EQVEAARKDLDOAKEEVRFLPQEEEGPGAGDESSCGTGGTHRRSKKAKAPERGCTRV 2283
1865 PRQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST---SSVPRPAATFPFATHCPL 1921
2284 ERLRGAEATQGANHTP-----VISAHQ-----TRSTTTTPRCSPARERVPREPAPRPT 2333
1922 GGTLDGVYPTLMEPVLLPKAEPVARPERPRADTGHAFKAPPARSGLEPASPSKSEP 1981
2334 PASAPAAIPAL-----VVPVSAVPVIGAPNPTITLPHILSPPPSPQIPCSSPA--CTP 2388
1982 RPLVPPVSGHATTIARTPAKNLAPHASP-----DPPAPPASAS-----DPHREKTSQKP 2030
2389 PPACTPPAHTP---PPAQTCLVTPSPPLLGPPSPVISAIVTNPLGLRPEALCAQAL 2445
2031 FSIQELERSLGVHSGS---YSPEGVSVS---PV-----SSPSLTHDKG-LPK 2072
2446 ASPSELEASVASSSETSSLSLVPKOLLPVAVEILPVSEKNLSLTPSAPSLTLEAGSIPN 2505
2073 HLELOKSHLEGRKQKQGVKLGGEAHL-----HLRPLPESQSSSPLLOTAPGV 2126
2506 GQEQEAPDSAGGTLTVLP-----EGEELPLCVSESGNLELPPSAASDEPLEAD 2558
2127 KGHQVVVTLAQHSIVITDYTRHHPQOL--SAPLAPLYSFGCASCPVLDRPPSDLYL 2185
2559 R-----TSEELTEAKTPTSSPEKQELVTAEVAAPTSSSATSP----- 2598
2186 PPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGEDGIEPVSPPEGMTEPGHRSR SAVYPLL 2245
2599 -----EGPSAPRPPR----- 2608
2246 YRDGEQTEPSRMGSKSPGNTSOPPA--FFSKLTESNMAVKKQKQKINKLNTNRNEPE 2303
2609 RRTSADVEIRGQGTGRG---QPGPKVURKLPGRILVTVVEEKELVORR----- 2655
2304 YNISQGTETFINPAITGTGLMYTRSOAQHASTNMGLEAIRKALMGKYDOWEESPPL 2363
2656 ---QORGAASLTVPVSET-----SASFGSPSV--RMSGP-----ESSPI 2692
2364 S-----ANAFNPLNASILPAAMP--ITADGRSDH 2392
2693 GGPCEAAPSLSLTPPQQPIARRHIELVGTGGSPENGDCALLAITPAVKRRRRGRPK 2752
2393 TLTSPPGGGKAKVGRSPSSKAKS---PARG-----LASGRPPSVSSH 2434
2753 KNRSPADAGRVDEAPSSTLKGTNGADVPGPETIIVADPVLEPOLIFGPQPLGPOVH 2812
2435 SEGDCNRRTPLTNRVWE---DRPSSAGSTPPFYNPLMLRQAQWAS-----PPPP 2482
2813 -----RPNPLLSVPKRRGRPPKARDLPIP-----GTISAGDGNSESTQPP 2856
2483 GLPAGSGLAGPHHAWDEBPKLLC 2507

Db 2857 PHPSPLTPL-----PPLVC 2871
RESULT 39
ABB62899
ID ABB62899 standard; protein; 3201 AA.
AC ABB62899;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 15489.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers BW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07002.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 15489; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3201 AA;
Query Match 3.6%; Score 473; DB 4; Length 3201;
Best Local Similarity 18.6%; Pred. NO. 4e-17;
Matches 573; Conservative 378; Mismatches 1132; Indels 994; Gaps 130;
QY 77 ERSQELHLPESHVSYPPELGKSEMEFIESKRPRELLPDLPLRPPSLATQPGASEDLT 136
Db 183 EKGVGQVFNPDSS---LPE-----PCVYVLPPEVW-----SPAHDKLP 217
QY 137 KORSITGKLEVPSPSPPHPTDPELELVP--PRLSKEELIQNMDRVREITMVQQISKLK 194
Db 218 KNYL-----RITKDEHYSFDLHVTNVVALAENTCAYDIDPID-----EAWRLYN 264
QY 195 KXQQQLLEEAAKPEPEKPVSPPPPIESKHSRLVQIYIDENRKKAAHRIEGLG----P 250
Db 265 SDRAQCGAPPINATQPERVIE--ELEVRQWEQIQVIL-----KLEGLGIEFDE 311
QY 251 QVELPLYNOPSTROYHE-----NIKNOA-----MRKKLLY 283
Db 312 NVICDVCRSP--DSEANEMVFCNDCNICVHOACYGITAIPSGQWLCTCSMGIKPCVLC 370

Db 2262 -PQVATTPNPTLTSPIDG--KMTYPAQLSSCAEAVVSMPTTPTVAT---AKD 2315
Qy 2026 TQSKPFSIOELRSLGYHSSYSPEGV-----EPVSVSPSLTHDKGLPKHLEELD 2078
Db 2316 SPKRTSVSGSNLSKKQTHKSPOLPOGSKPGKPRQPLQPTTPAPVPVVALPPTKYDQ 2375
Qy 2079 KSHLEGELRPKQP-----GPKVLGGEAAHLPHLRPLPESQPSPL- 2120
Db 2376 THTLOKPRQAPRGSGGAPGRGRGRGGGVTSGMAWLPVPPMDSYGSNTHV 2435
Qy 2121 -----QTAPGVKHQVRVTVLAOHISEVITQDTRHHPOOLSAPLAPLY 2164
Db 2436 NNIVGTPFENFEDDMAGGVENLQSL-----RDRRSFELRAP---RVQ 2478
Qy 2165 SFGASCPLVDLRPPSDLYLP--PPDHGAPARG--SPHSBGGKRSPEPNKTSVLGGEDG 2221
Db 2479 NKETTTTATTNPLLHPVLPQVDMRTYNLGFAPHSTA---SQEAYQNLLGAFDSG 2535
Qy 2222 IEPVSPPEGTEFGHRSASVYLLYRDGEQTEPSRMGS-----KSPQNTSQPPAFPS 2273
Db 2536 TADQTLSEFNEEDERQFOSA---LRATGTGTSPSKQHSQPTALVAPPTGPNTPAPNLLL 2592
Qy 2274 KLTESN-----SAMVSKKQEI-----NKLNTHRNPEP 2302
Db 2593 HCTEANQMAPNVAATGAATHLVEGSLVEASLEATSEVSDSDSTIPHSKTSSTDARSQI 2652
Qy 2303 EYNISOP-----GTIFENPAITGTGLMTYRQAOVQEHASTNMGLEAIRKALMGKYDQ 2356
Db 2653 KLKIKFPMAYPEHYNAMNTSSSITLSTLVQSSNVQTTVSTTVSA--SSAVGNSRR 2710
Qy 2357 WEESPPLSANAFNPLN-----ASASLPAAMPITAAG--RSDHTLTSPGGGKAKVSGRPS 2410
Db 2711 MRKELLSLVVQKNDHNDSSCGLPAASDTLPLENLKSEERDELGGNGTKRFKNSS 2770
Qy 2411 SRKAKSPAPCLA-----SGD-RPPSVSVHSEGDGNCNRTPLTRWE 2451
Db 2771 SRELALDANLALVEEQLLSSGACGGSGGRRRSACSGSNDNNGKTAASAGK 2830
Qy 2452 DR-----PSSAGSTPPFNPPLMRLQAGVWAGPPPPGLPAGGGLAGPHAWDE----- 2500
Db 2831 RRGSKTLESSEDDHQAOKLIKIR---GLTANETSGV---SSVDEGQNSYEMTERRAC 2884
Qy 2501 EPKPLLCQSYETLSDSE 2517.
Db 2885 PPKRLTNSFTLTLEE 2901
RESULT 40
ABR92087
AC ABR92087 standard; protein; 2701 AA.
AC ABR92087;
XX
XX
XX 10-SEP-2003 (first entry)
DT Human cervical cancer cell marker encoding cDNA SEQ ID NO:83.
XX
XX DE Human cervical cancer; cervical cancer marker; cancer therapy;
KW Human; cervical cancer; cervical cancer marker; cancer therapy;
XX detection; gene therapy; vaccine.
XX Homo sapiens.
XX WO2002101075-A2.
XX
PD 19-DEC-2002.
XX
PF 12-JUN-2002; 2002WO-US018638.
XX
PR 13-JUN-2001; 2001US-0298155P.
PR 13-JUN-2001; 2001US-0298159P.
PR 14-NOV-2001; 2001US-0335936P.
XX

PA (MILL-) MILLENNIUM PHARM INC.
PI Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;
PI Gannavarapu M, Glatt K, Hoersch S;
XX
XX WPI: 2003-156967/15.
DR N-PSDB; ACF12869.
XX
PT New isolated nucleic acid molecule useful for detecting, characterizing,
PT preventing and treating human cervical cancers, in various prognostic and
PT diagnostic assays, in pharmacogenomics and in monitoring clinical trials.
XX
XX Claim 4; Page 233-239; 386pp; English.
XX
CC ACF12828 to ACF12947 encode the human cervical cancer marker proteins (I)
CC given in ABR92047 to ABR92164. A higher level of expression of (I) than
CC normal indicates the presence of cervical cancer. Also described: (1) a
CC vector (II) containing (I); (2) a host cell (III) containing (I); and (3)
CC assessing (MI) whether a patient is afflicted with cervical cancer,
CC comprising comparing the level of expression of a marker in a patient's
CC sample, and the normal level of expression of the marker in a control non
CC -cervical cancer sample, where a significant increase in the level of
CC expression of the marker in the patient's sample relative to that in the
CC control sample is an indication that the patient is afflicted with
CC cervical cancer. (I) has cytostatic activity, and can be used in gene
CC therapy and in vaccines. (I) is useful in detecting, characterizing,
CC preventing and treating human cervical cancers. (I) may also be used in
CC various prognostic and diagnostic assays, pharmacogenomics and in
CC monitoring clinical trials
XX
SQ Sequence 2701 AA;
Query Match 3.6%; Score 469.5; DB 6; Length 2701;
Best Local Similarity 19.1%; Pred. No. 5e-17;
Matches 532; Conservative 332; Mismatches 980; Indels 944; Gaps 130;
Qy 163 VPRLSKBELIQMDRVREITWVEQOIKLKKQOOLEEEAAK--PPEPEKP--VSPPP 218
Db 54 MPPANPLSLKAEKNGDNPVNIKPDGTGWSKQBEKEETPEVPPAQKPGVAAPPE 113
Qy 219 IESKHSLSQIIVDENKKAEEAHRILEGLGPOV-----ELPLXNQPSDTRVH----- 267
Db 114 VAPAKS-----WASNKQG-----QGDGIQVNSQFOEFPSLQAAGQEKKEKETND 161
Qy 268 ENIKINQAMRKLLILYFKERNHA-----RKWKQKF-----CORYDQLEALEKKV 313
Db 162 DNVGPGSLRPPNVACWRDGGKAAGSPSSDQDEKLPQDSEAGTSEQNDILKVKVKRI 221
Qy 314 -----ERIENPRRRAKESKVR-----EYBEKQFPEIR-----KORELQERMSRVRGQS 359
Db 222 ACGPPQAKLNGQQAALASQVAMMPYMFQYPRMTYPLHGPMPRPPSLSETNKGILGR 281
Qy 360 GLSMSAARSEHEVSEIIDLSEENLEK-----OMRQLAVIPMLYDADQ 404
Db 282 GPPPSWA--SEPERPSILSA--SELKELDKFDNLDAADEGWAGAQMEVDYTEQLNFSDDDE 339
Qy 405 QRIKFTNMGLMADPMKYKQVDMVMWSEQKET-----NENKETDEVSNKSSQIPAPSVAKVPYK 439
Db 340 QGSNSPKENNSDQGSKASEN-----FMQHPKNF-----GLIASFLERKTV--ECVLYYYLTKN 392
Qy 440 ---FREK-----FMQHPKNF-----GLIASFLERKTV--ECVLYYYLTKN 476
Db 393 GPSFNQERTGSSHLPPPPKLLAQHPPDRQAVPGRPPGPPPSKQVDADEIWKQRRRQ 452
Qy 477 ENYKSLVRSYRRGKSKSQOQQOQQOQQOQQO-----QQQPMPSSEKDEKE 524
Db 453 SEISAVERARKEREEREERMEEQRKACAEKLRLEKLGILEKQSPSEIRER---ERE 510
Qy 525 KEKEASEKEKEPEVNDKEDLLKEKTDTSGEDNDEKEAVASKRGTANSQGRKGRITR 584
Db 511 KEREREKELEK--EQEERE---KEREKDRQOKEKEKEKEKEKO-----R 553
Qy 585 SMANEANSEEAITPOOSAEIASMELNNESSRWTEEMETAKKGLLEHGRNWSAIRMVGSK 644

Db 554 EMEKERKQEK-----EKELEKQK----- 571
QY 645 TVSQCKNFYKKNQNLDEILQOHK-LKWEKERNARRKKKAPAAASEEAAFPVVEDE 703
Db 572 -----EKELEKQEKQEKQEKLEKLEK-----IEPRENLEPMVEKQ 614
QY 704 EMEASGVSGNEEMVEEAEALHASGNEVPRGECGSPATVNGSDTESIPSPHTEAAKDTG 763
Db 615 ESENS--CNKEBEPVTRQD-----SNRSEKEATPVVH-ETEPESS 652
QY 764 QNGKXP-----PATLGADGPPPPGPT 784
Db 653 SQ-PRPAVLGKFKQKSLPRFORQOMKQOQOQOQOQVLPQTVPSPSSSTVPP 711
QY 785 PRRRTSRAPETPASEA-----TGAPTP--PRAPPSAPPVV 822
Db 712 PHRPPLYQMPHQHSLASMGFDPRLMMQSYNDPRMMSGRFAMDIPPIHPGMIIPKPLM 771
QY 823 PKEEKEETAAAPPVEE-----GEEKPPAABELAVDTCKAEEP-- 861
Db 772 RDQMEGSPNSSEFHIARSARDHAI SLSEPRMLWGSOPYPHABPQOATTPKATEEPED 831
QY 862 VKSECTEAEAGPAKQDAEAEATAGALKAEKKEGGGRATTAKSGAPO----- 913
Db 832 VRSEALDQEQ-----ITAAYSVEHNQLEAHPK-----ADFTRESSEAQVOKFLSRS 878
QY 914 -----DSOSSATCSADEVDEAGGDKNLLSPRPSLLTPT-GDPRANAS----- 956
Db 879 VEDVRPHHTDANNQSA-----FEAPDQKTLASAPQERI SAVESQPSRKRSVSHGNSH 931
QY 957 POKPLDLKOLKORAAAIPI-----QVTKVHEPPREDA-----APTKEPAPP----- 997
Db 932 TQXPDE--QRESPSAGIPKVTSCIDSKEPIERPEEKPKKEGFIRESSEKPEKVKYKKS 989
QY 998 -----APPPONLOPE--SDAPQO--PGSSPRGKRSPPAPPAKEAFAPAAQKLPDGPDPCH 1049
Db 990 ETRWGRPSNRREEVNDRPVRRSGPIKKPVLDMKEEREQRKEGEKAKE----- 1041
QY 1050 TSGCLPFPVPREVIKASPHAPDPSAFSYPAGHPLPLGLLHDHTARVLPRLPPTISNPPPLI 1109
Db 1042 -----VTEKVVK-----PEKTEKKOLPPPPPPP-----QPPAPIQOQ-SVPPPIQ 1081
QY 1110 SSA-KHPSVLEROIGAISQMSVOLHVPYSEHAKAPVGVMTGLPLPMDPKKLAPPVGVK 1168
Db 1082 PEAEKEPST--ETATLAKQS-----QTEK--PLEPVS 1111
QY 1169 QBQLSPRQAGPPESLGVTQAQASVLRGTALGSPVGGSIKGIPTRVPSDSAITYRS 1228
Db 1112 TVQVEPAVKTVNOQTMAAPVVKEE-----KQPEKVISKOLVIERPRDPSRPVAKKE 1162
QY 1229 ITHGTPADVLTKTITRIIGESPSRLDRGSDLPKGHVI--YEGK--KGHVLSYE 1281
Db 1163 ST--LPRTVTKBARERDWFDDQYR-GRGGBEYYSRGRSYRSGYGRGRGGRGHTRDY- 1218
QY 1282 GGMSTVQCSKEDGRSS--SGP-----PHETAAPKRTYDMMEGRVGRAS----- 1323
Db 1219 -----POYRDNKPRAEHIISGFLRQRESESTRSESSDFEVPVKRRQRGSETDTSIEHE 1273
QY 1324 SASIEGLMGRAI PPERH-----SPH-HLKEQHIRGSIQTQIPRSYV-----BAQED 1369
Db 1274 SASDKDLSKGLPKREERPENKPKVKPHSSFXPDNHVRID-NRLLEKPYVRDDDKAKPG 1332
QY 1370 YL-----RREAKLKRECTPPPPPSRDLT--EAYKQALGPKLKPAPHEGLVATVK 1419
Db 1333 FLPKGBPTARGGTFRRGDRDPGGRPSRPSLURRPAYRDNQNNPQO----- 1379
QY 1420 EAGRSIHEIPREB-----LRHTPELPLAPRLKEGSIQTQPLKYDTGASTGSKKHVDR 1474
Db 1380 -----SEVPKPEDGEPRRHEQIFIA-----ADKRPKPE-----RKFD-- 1414
QY 1475 SLIGSGRTFP-----PVHP--LDVMADARALERACYEESLKSPPGTASSSGSGIARGA-- 1526

Db 1415 -----PARERPRRQRPTRPPRODKPFRRLRER--EAASKSNEVVAVTNGTVNNVAQE 1467
QY 1527 PV-----TVPELGKPRQSLTVED-----HGAPFAGHLPRG-----SPVTM 1562
Db 1468 PVTNLGDISGNTKTPDLSNQNSSDAQNEWETASESDFNERRERDEKKNADLNATQTVVKV 1527
QY 1563 RE-----PTPRLQESLSSKA--SQDRKLTSTPREIAKSPHSTVPEHHPHPTIS-----P 1610
Db 1528 GENVLPPKREIAKRSFSSQRPVDRNRGNNGPPKSGRNFSGPRNERRSGPPSKSGKRGF 1587
QY 1611 YEHLLRGVSGVDLYRSHIPLAFDPTISIPRGIPLDAAYLPHLAPNTPHYLPPYLI 1670
Db 1588 FDDQFAGTIGDVL-----INGSAHHQ----- 1609
QY 1671 RGYPTAALENRQTIINDYITSQMHNTATAMQADMLRG-LGPRESSAL-----N 1723
Db 1610 EGVNGTGCKNSKD-----STGKREDPKPKPKPKKEKVDALSQFDLNN 1653
QY 1724 YAGPRGIIDLSQVPHLPVLVPTTGTATAMDRILAYLTAPOPFSSRSHSSPLSPGGPT 1783
Db 1654 YAS-----VVIIDHPEVTIEDP-----QSNLNDGGT 1682
QY 1784 HLTKTPTTSSSERDRDRDREREKSIILSTTTVEHAPIWRPCTQSSGSSSGG 1843
Db 1683 EV-----VSKOOKLQDEER--RKKEQVI-----QVMNKNANEXGRS----- 1720
QY 1844 GGGSSSRPASHAHQHSPIPR-----TQDALQQRPSVLHNTGMKGIITAVPSPKPTVLR 1899
Db 1721 -----QTSKLPFRFAKQATGQQAQSS--SASVPPPLASAPLPPSTASV 1762
QY 1900 STSTSPVRPAATFPATHCPGLGGLDGYVP-TLMEPVLLPKAPRVARPERPRADTGH 1958
Db 1763 PASTSAPL-----PATLTPVASTSAPVASTLAPLASTSAPVPASPLAPVSASV 1815
QY 1959 FLAKPPARS--GLEPASSPSKSGSERPLVPVSGHATI-----ARTPAKN 2001
Db 1816 SASVFASTSAAAITSSSAPASAPATPILASVSTPASVTILASAPIILASALASTAPT 1875
QY 2002 LAPHASPPPP--APPASASDPHREKTSQKSPFSIQEELRSILGVHSGSVSPGVPVSP 2058
Db 1876 PAPAASSPAAPVITAPTIPASAP-----TASVELA-----PASASAPAPAP-TP 1918
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGELPKQPGPVKLGGEAAHLPHLPLPESQPSSS- 2117
Db 1919 VSAPN-----PAPPAQAQOQ-THKPVQNPLOTTSSSQK 1953
QY 2118 -----PLLOTPAGVKGHQVRVTLAQHISEVITQDYTRHHPQO--LSAPL-----P 2160
Db 1954 PPSIRLPSAQTPNG-----TDVASGKSIQTPQSHGTLTAEMLDNKVAPP 1999
QY 2161 APLYSPPGASCPLDLRRPPSDLYLPPDPHGAPAR-----GSPHSEGGKRSPEPNKTSV 2214
Db 2000 AVLNDISKKLGI-----SPPO-----PPSVSAWNKPLTSGSAPSSEKAKNGQE--SG 2046
QY 2215 LGGEDGIEBPVPEGMTEPGHRSNAVYLLYRDGEQ--TEPSRMGSKSPGNTSQPAPF 2272
Db 2047 LEIGTDTIOFGAPASNGNE-----NEWPVLSEKADKIPPEKQKQKQF--RAGPIKA 2098
QY 2273 SKLTESNAWVSKQO-EINKLINTHNRN-----EPEVN----- 2305
Db 2099 OKLPDLSPVENKEHKGPIGKERSLKNRKVKQDAQOQVEGQEKPSPATVRSPTDPTTKET 2158
QY 2306 --ISQPGTIFNPAITGTGLMTYRSQAOVEHASTMNGLEAIRKALMGKYDQWESPPL 2363
Db 2159 KAVSEMSITGIMI SVSSAEYGTNAKESVTDYTPSSSLPNTVATNTKMDTLVNNVPL 2218
QY 2364 SANAFNP-----LNASALPAAMPITAADGSRSHLTSPCGGGKAKVSGRPSRKAKSP-- 2417
Db 2219 PNTPLPKRETIQSSSLTSVPPTT-----FSLTFKMSARKAWENSPNREKSGSPVT 2271
QY 2418 -----APGLASODRPPSVSSVHSEGDGCMR-RTPLNVRWEDRPPSAGSTPPPNPLMR 2470
Db 2272 STAPPIATGVSSASGSPSTANYNSFSSAMPQIPVASVTPTASLSAGT-----YTTSSLS 2327

Qy	1282	GHMSVTCQSKEDGRSS----	SGP----	PHETAAPKRTYDMMEGRVGRAIS-----	1322	
Db	1221	-----POYDNKPRAEHIPSGLPQRESESTRESSDFEVPVRRRRQRGSETDTSDEIHE	1275	-----		
Qy	1324	SASTIEGLMGRALPPERH-----	SPH-HLKEQHIIIRSGSIITQGIIPRSVV-----	BAQED	1369	
Db	1276	SASDKOSLSKGKUPKREERENKPKVPKPHSGSPDNHVRID-NRLLEKPIVRODDDKAKPG	1334	-----		
Qy	1370	YL-----BREAULKKREGTPPPPPPSRDLT---	EAYKTQALGQPLKPKLPAHEGLVATVK	1419		
Db	1335	FLPKGEFTRRGCGTFRGGREDGGRSPRSRSTLRRPAYDRQWNPRO-----	1381	-----		
Qy	1420	EAGRSIHEIPREE-----	LRHTPELPIAPRLKEGSITQGTPTLKYDTGASTTGSKKKHVDR	1474		
Db	1382	-----SEVPKPEDGEPRRRHEQFIPIA-----	ADKRPPKFE-----	RKFD--	1416	
Qy	1475	SLIGSPGRTFP-----	PVHP--LDVMADARALACRYEESIKSRPGTASSGGSGIARGA--	1526		
Db	1417	-----PARERPRQRTPRPPQDKPPRFLRER--	EAAKSNEVVAVPTNGTVNNVAQE	1469		
Qy	1527	PV-----	IVPELCKPRQSPLTVED-----	HGAPFAGHLPRG-----	SPVTM	1562
Db	1470	PVNTLGDISGNKTPDLNQNSSDOANEWETASESSDFNERERDEKKNADINAQTUVKV	1529	-----		
Qy	1563	RE-----	PTPRLQBSLSSSKA-----	SQDRKLSTPREIAKSPHSTVPEHHPHPIIS-----	P	1610
Db	1530	GENVLPKREIAKGSFSSQRPVDRQNRGNGPPKSGRNFSGPRNRRSGPPSKSGKRGP	1589	-----		
Qy	1611	YEHLLRGVSGVDLYRSHIPLAFDPTSI	PRGIPLDAAAYYLPHLAPNPYPHLYPPYLI	1670		
Db	1590	FDDQPACTTGVDL-----	-----	INGSSAHQO-----	1611	
Qy	1671	RGVPDTAALENROTIINDYITTSQMHENTATAMAQRADMLRG--	LSPRESSLAL-----	N	1723	
Db	1612	EGVPNGTGQKNKD-----	-----	STGKKREDPKPGPKPEKVDALSQFDLNN	1655	
Qy	1724	YAAGPRGIIIDLQVPHPLVLPVPTPGTATAMDRLAYLPTAPOFFSSHRHSSSPLSPGGPT	1783	-----		
Db	1656	YAS-----	VVIIDHPEVTVEDP-----	-----	QSNLNDGGFT	1684
Qy	1784	HLTKPTTSSERDRDRDRDRDRREKESILSTTTTVEHAPIWRPGTEOSSSSSSSGG	1843	-----		
Db	1685	EV-----	VSKQQRKLQDEER--	RKKEEQVI-----	QWNKKNANEKGRS-----	1722
Qy	1844	GGGSSSRPASHAHQHSPISPR---	TQDALQORPSVLHNTGMKIITAVEPSPKTVLR	1899		
Db	1723	-----	QSKUPPRFAKQATGIIQAQS-----	SASVPPPLASAPLPPSTASV	1764	
Qy	1900	STSTSPVRPAATPPPATHCPGLGTGDGVYP-TIMEPVLPLKEAPRVARPERPRADTGHA	1958	-----		
Db	1765	PASTSAPL-----	PATLTPVPASTSAPVPASTLAPVLASTSAPVPASPLAPVPSASASV	1817		
Qy	1959	FLAKPPARS--	GLEPASSPKSGSPRLVPVPSGHATI--	-----	ARTPAKN	2001
Db	1818	SASVPASTSAAAITSSAPASAPAPTILASVSTPASVTILASASIPILASALASTSAPT	1877	-----		
Qy	2002	LAPHHASPDP--	APPASADPHREKTQSPFSIQLELRSLGYHSGSSYPEGVEVPSP	2058		
Db	1878	PAPASFPAPAVITAPTITPASAP--	TASVPLA-----	PASASAPAPAP-TP	1920	
Qy	2059	VSSPSLTHDKGLPKHLEBDKSHLEGBLRPKQPGFVKLGGEAAHLPHRLPLPSPQSSS-	2117	-----		
Db	1921	VSAFN-----	PAPPAPAQTOAQ-THKPQNPLQTTSSQSKQ	1955		
Qy	2118	-----	PLIQTAPGVKGHORVTLQAHISEVITQDYTRIHPQO--	LSAPL-----	P	2160
Db	1956	PPPSIRLFSAQTPNG-----	-----	TDYVASGKSIQTPOSHGTLTIAELMDNKNVAPP	2001	
Qy	2161	APLYSFFGASCFVLDLRPPSDLYLPDPDHGAPAR-----	GSPHSGGGKRSPEPNKTSV	2214		
Db	2002	AVLNDISKULGI-----	SPDQ-----	PPSVSANNKPLTSGFASPSSGAKNGQB-----	SG	2048
Qy	2215	LGGEDGIEFPVSPPEGMTEFCHRSRVSVPVLYLRDGEQ--	TEPSRMGSKSGCNTSQQPPAFF	2272		

Db	2049	LEIGTDTIOGAPASNGNE-----NEWVPVLSSEKADKIPEPKQKQKP---RAGPIKA 2100
Qy	2273	SKLTESNSAMVSKKQ-EINKKLNTNHN-----EPEYN----- 2305
Db	2101	OKLPDLSPVENKEHKPGICKERSLKNRVKDAQQVEPEGQEKPSPATVRSSTDPTTKET 2160
Qy	2306	--ISQPGTEIFNMPAITGTGLMTYRSQAOEHASTNMGLEAIIRKALMGKYDQWESPPPL 2363
Db	2161	KAVSEMSTEIGTMISVSSABEYGTNAKESVTDYTTSPSSLPNTVATNNTKMEDTLVNNVPL 2220
Qy	2364	SANAFNP-----LNASASLPAAMPITAADGRSDHILTSPGGGKAKVSGRPSSRKAKSP-- 2417
Db	2221	PNTLPLPKRETIQQSSSLTSVPETT-----FSLTFKMSARSAKAVENSPNVREKSPVT 2273
Qy	2418	-----APGLASGDRPPSVSVHSEGDGCR-RUPLNLRVWEDRPSGAGSPFPYVNPILMR 2470
Db	2274	STAPPTATGVSSASGFSSTANYNSFSSASMPQIPVASVTFPTASLSGAGT----YTTSSLS 2329
Qy	2471	LQAGVMASPP-----PPGLPAGSGPLA 2492
Db	2330	TKSTTTSDPPNICVKVPQQLQTSLSLPSA 2357
RESULT 42		
AB	BB71160	ID
AB	BB71160	standard; protein; 5560 AA.
AC	BB71160;	
XX		
XX		
DT	26-MAR-2002	(first entry)
XX		
XX		
DE	Drosophila melanogaster	polypeptide SEQ ID NO 40272.
XX		
KW	Drosophila;	developmental biology; cell signalling; insecticide;
KW	pharmaceutical.	
XX		
OS	Drosophila melanogaster.	
XX		
PN	WC200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	23-MAR-2001;	2001WO-US009231.
XX		
PR	23-MAR-2000;	2000US-0191637P.
PR	11-JUL-2000;	2000US-00614150.
XX		
PA	(PEKE)	PE CORP NY.
XX		
PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX		
DR	WPI;	2001-656860/75.
DR	N-PSDB;	ABL15263.
XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	genes from Drosophila and for elucidating cell signaling and cell-cell	
PT	interactions.	
XX		
PS	Disclosure; SEQ ID NO 40272; 2lpp + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent	
CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
CC	useful in developmental biology and in elucidating cell signalling and	
CC	cell-cell interactions in higher eukaryotes for the development of	
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	discloses genomic DNA sequences (AB116176-AB130511), expressed DNA	
CC	sequences (AB11840-AB116175) and the encoded proteins (ABBS7737-	
CC	ABBS72072). The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX		
SQ	Sequence 5560 AA;	


```
Db 4187 VDSYWSAKEVNIDSVIKKLDTPLA--SKDAKRAVEMQAIAAPAPFN---PQFGNSMAQ 4241
QY 1667 PYLIRGYPTAALENRQTIINDYITISQOMHNTATAMAQRA-----DMLR 1711
Db 4242 ETAL-----PTTSMVNS---NDHDEDE-----TETROLPPAKPIPTVGRPPRGSGSAKR 4291
QY 1712 GLSPRES-----SLALN-YAAGRGIIDLSQVP-----HLPVLVPPPTGTPA----- 1752
Db 4292 GRQPRGAKKVGFPPLNSVTAAPGVDSLVPQDNGVQVTLKPVTAFTVTRGKGRPPRN 4351
QY 1753 -----TAMDRLAYLPTAQPFSSSHSSPLSPGGPHTLTKPTTSSSERDRDRDRD 1805
Db 4352 LLLQQOOLQQQLIDIRKGMEMVTSATSTPL-----PTDIPTSSVLTAAEK-KARNAQLT 4406
QY 1806 RDREREKSLTSTTVEHAPIWPCPEQSSGSGSGSGSGSSRP-----ASHSHAH 1858
Db 4407 QAOEQNQVASQVGTGDIYEFHEDGGEPPKPTISSVAPSABDQRPRLTLTINKTQPSIK 4466
QY 1859 QHSPISPRQDALQORPSVLHNTG-----MKGIITAVEPSKPTVLR 1899
Db 4467 NISEMEQTIQQOQQOQSEVISNTDPIDGDNBSBSCNTRKSRRLQEKEDRSTVDDIIEDVVR 4526
QY 1900 STSTSPVRPAATFPFATHCPGLGTLGDVYPTLMPEVLLPKEAPR----- 1944
Db 4527 NTNT-----PTGTGP---HLPKGAQTP-----PRSGRNAQAQKTDVAQIINA 4566
QY 1945 VARPERPR-----ADTGHAFIAPKPARSGLE----- 1970
Db 4567 VGRPRSKDKRTIGETANTLIEBVTASNATVAASHLA-PPEGAGVESHVPDQAKEVEPV 4625
QY 1971 ----PASSPSKGEPRPLVPVPSGHATIAARTAKNLAPHAGPDPPAPPASADPHREKT 2026
Db 4626 SVVTPISTTAPSVAAPTVFPV-----AMVPVKPTMPOH--PKKALIAAAIESQAIN 4678
QY 2027 QSKPF-----SIOBLELRSLGYHSGSYSPGVEVPVSPVSPSLTHDK-----GL 2070
Db 4679 SSTPSGGLPMHQTAAATQKITGGVADAVSKALVDVPTGITAGMPQGEKNLPAATAAA 4738
QY 2071 PKHLELD-----KSHLEGELRPKP--GPVKLGGEAAHLPLRLPESQSPSSPLL 2120
Db 4739 PANSNEDQOAPPPOLOHQOQQOQHPPQOQOQANLQINTLIPSLPNTIATLGKSVQL 4798
QY 2121 QTAPGVKHQVRVTLAQ-HISEVITQDYTR-----HHPQQLSAPL 2159
Db 4799 ETSAAALLNKPSVLKGNASVIOQQQPIQIVAPAKPIILQONPLPTVLHQAHTTVRP 4858
QY 2160 PAPLYSFPGASCPVLDLRR-----PPSDLYLPPDPHCAPARGSPHS-----EGGK 2204
Db 4859 POPL-----KARVLNREKNIQQQLTPTKQAVAPQPOH-----APHSGHMLLTDTAGNQ 4906
QY 2205 RSPERNKTVLGGEDGIEPVSPPEGMTPEGHSR-----SAVYPLLVRDGEQTE 2253
Db 4907 QLVQPOILIAHHQQOQHLOQNVNPPPTTAHSPHSRIPISQOQQQLGPGASISP-----QOQQ 4960
QY 2254 PSRMGSKSPGNTSQPPAFFSKLTESNAMVSKKQBEINKLANTHNRNEPEYNISQPTBI 2313
Db 4961 PQTIVTIKQASAAQPOIL--HVVSRSKASVVPQOQO---QLPPTSSTGHLQLAKENYSY 5015
QY 2314 F-----NMPATGTGLM--TYRSQAVQEHASTNWGLEAIRKA-----LMGK 2353
Db 5016 APTVLPTPLFAVQOQQOQHLYKONNQKGAQIQMPHPHGIIMPTHPCGMLLQOKLPAHLPOQ 5075
QY 2354 YQOWESSPP-----LSANAFNPLNASLASLPAAMPITADGRSDHTLTPSGGGKA 2403
Db 5076 QHQLNPSPPGKNPNVLHGLQSQIMPGSVGSPPPVSAAVLTKTAQQOVNSVVPVAGIRTA 5135
QY 2404 KVSGRPSRSRAKSP---APCLASGDRPPSVSVHSEGD---CNRRT-----PLTN 2447
Db 5136 IPNISPQSPRVSPVLVPPGISG--VPPPDASLHDLGAVSGRRTQSPPPPAHQOASPIIP 5193
QY 2448 RWEDRPSAGSTPPFPNPLIMRLQAG-----VNASPPPPGL--PAGSGP---LAGPHAWD 2499
```

```
Db 15194 NDSTYRGVTASRDFMLYQHHLMR--GGDYDDKMGSSPPLRLRRPGSGPRTTIAVPH----- 5247
QY 2500 BEPKPLLCQYETLSDS 2516
Db 5248 -----SLOSPQDRTAADS 5260
RESULT 43
ABB60327
ID ABB60327 standard; protein; 3111 AA.
XX
AC ABB60327;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7773.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04430.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX
PS Disclosure; SEQ ID NO 7773; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3111 AA;
Query Match 3.5%; Score 458.5; DB 4; Length 3111;
Best Local Similarity 18.1%; Pred. No. 2.5e-16;
Matches 534; Conservative 376; Mismatches 1061; Indels 975; Gaps 123;
QY 3 GSTQLVAQVATRATEPRYPHSLSYFVQIARTHTDVGLEYYQHHSRDYASHLSPG---SII 59
Db 553 GETLL---WRDTPD-----SRRRRD--SLTWKETKADRAAMIREGVDVSV 593
QY 60 QQRRR-----PSLLSEFQCNERS-----QELHLRPESHVYL 92
Db 594 RTQKLRTFRGKSDSSYSDSDSGEQSGTGCGGGSTDTSLCDDDDPKSTKSPKQAKL 653
QY 93 PELGKSEMFIEFKRPRLELLPDLLRPS---PILATQGPAGSE---DLTKDRSLTGKL 145
Db 654 ARKLKEQKQLAGSRSTSLERQRPKSWAPSSHPIFPMLMGTDSGDEKEDSKTEGPIGDO 713
```


Db 2634 L--RTGENTINDRUSSIKNS-IKSIDTLCE-EKPYQKEKQRYIDSLFTSLHFAKSSSL 2689
QY 1896 -----TVLRSTSTSSVPRPAATEPP-----ATHCPLGTLGTVVPTLMPEVLLKPEAP 1943
Db 2690 EDLSLSRSLSRSEGRGSHRSYDAPSRVSEHRSLS-GSADS-----RRSP 2736
QY 1944 RVARPERPRADTGHAFKAK--PPARSGLPFAPSPSKGSEPRPLVPVPPVSGHATIAATPAKN 2001
Db 2737 LGNRDTSPLHRRSHRDISRELSPRRLLEEDEERKDRS-----SVVRDN 2783
QY 2002 LAPHPASPPPPAPASAS-----DPHREKTQSK-----PFSIOELELRSLG 2042
Db 2784 LLFYPADNRSELSSGSLTGFNHVKVDRQLEETCAKYADRRSACRTPLS-HPYESRTTA 2842
QY 2043 YHGSSVSPGVEPVSPVSPSLTHD--KGLPKHLELDKSHLEGELRPQGPVKLGGEA 2100
Db 2843 TRHSHTDPVQI-PTNPAGSATATDSPPRPVSPRYQPYDPH-----RSPGGA--GGTP 2892
QY 2101 AHLPL-----HLRPLPESQPSSSPLLQATAPGVKGHRV 2132
Db 2893 LYQPGKLEIRHTTSTFYDRFLTEKQIERQTHSRP-----PFRSPVSPVPAKSYVEL 2947
QY 2133 VTLAQHISEVITQYTRHHPOQLSAPLAPLYSPGASCPLDLRRPPSDLYLPPDHGA 2192
Db 2948 CSTSGTSTATSTSTSSSF---MSSSYAGFSFSLPSAS-----NFSYLNLP----- 2989
QY 2193 PARGPSHSGGKSPENKTSVLGGCGDIEPVSPPEGMTEPCHRSVAVYLLYRDGEOT 2252
Db 2990 ---GS-----GSGSGISSISP----- 3002
QY 2253 EFSRMGSKPGNTSQPPAPFSLKTESNANVSKKQEIKNKLNTHNRNPEYNIQPGTE 2312
Db 3003 ---RASCSDLRSTSGTSTSTSTVTSYV-----PYNFTSSFTS 3040
QY 2313 IFNMPAITGTGLMYRSQAVQEHASTNMGLEAIRKALMGKYDQWEESSPLSANAFNPLN 2372
Db 3041 RLNDPIITSTSAVSTSSLTSTHSTGVYNNPMWSEFTLRE-----PLASSSIGSS 3087
QY 2373 ASASLP 2378
Db 3088 ASPLLP 3093

RESULT 44
ID ABB65772
XX ABB65772 standard; protein; 5533 AA.
AC ABB65772;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 24108.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacological.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US0009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers BW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL09875.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutic and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5533 AA;
Query Match 3.5%; Score 458.5; DB 4; Length 5533;
Best Local Similarity 17.9%; Pred. No. 5.4e-16;
Matches 602; Conservative 370; Mismatches 1047; Indels 1337; Gaps 147;
QY 31 ARTHTDVGLLEYQHHSRDYASHLSPGSIQPORRRPSSLSEFPQGNERSQELHLPESH 90
Db 2358 AATNTSSGKLHHQHRVERKSRG-----SDEGH-----HS 2390
QY 91 YLPELGKSEMEPIESKRPRLELLPDLRPSPLATGPAGSEDTKORSLTGKLEVPSP 150
Db 2391 -----SSKSLRAKLM-----MLSSADSDTDDASKHSIF----- 2420
QY 151 PSPHTDPELELVPRLSKEELIQNMDRVDRREITWVEQIQISKLKKQOOLEEAAKPPPP 210
Db 2421 -DIPDDCNVSMYDK--VKARSKNMQOAE-----KKAKFQOLKOSRAK----- 2465
QY 211 EKPVPSPPIESKHSVLQIYD-----ENRKAAEAHRILEGIGQVPELP----- 255
Db 2466 -----KCRS---TSYDGDSDTEFEDQHRNSGSSPHGRYPGLSSDDDDDBET 2511
QY 256 -----LYNQPSDT-----ROYHENIK-----INQAMKKLI--LY 283
Db 2512 HORRISSDAEHGGDQNGASTLADANVRQMNRLCDGDDSDSEDEIRNVMKHS 2571
QY 284 FKRRNH-----ARKQW---KQKFCQYDQLMW 307
Db 2572 FGRNNSNTIASDSESQSPAPDLTIQEHPIAPAQEIKREQLSDEBQKFSRHSNNS 2631
QY 308 ALEKKVERIEN-----NPRRAKESKVREYKEOFFEIRKO-RELQERMQRVGOR 357
Db 2632 IERKLTERTKTELGDYFNSSEYTYTGKLEY---SPETRKHKHKKSLKS----- 2682
QY 358 SGLSMAARSE-----HEVSEI-----IDGL-----SE 381
Db 2683 SSTADTSAATPLVMTPLTSPIDFVHSSECKTKFDNFDLKTCSIPLEISAGERAKH 2742
QY 382 QENLEKQKROL-----AVIP--PMYDADQORI-----KFTN 411
Db 2743 KERKKEKRLNMTTEATVPNSPTNTDSSEKLSKEERHRLKSKSKMSMNSCNTKIYN 2802
QY 412 MNG-----IMADPMKVYKDRQ-----VMNMWSEQEKETFREKFMQHPQNF 452
Db 2803 SSGAHPSTSPSLPATPTSAPTAQTAKRGEDKMEFI FGLISDEESQFPE---QAETNKD 2859
QY 453 LIASFLEK-TVAECVLYYY-----LTKKNVKSIVRRSRRRGSKSQOQQOQQOQQO 507
Db 2860 IIPSSVSTTGPVSAALQTYQEPSTPNKNEEAHIQLTV-----HEPEQQOQLERS 2912
QY 508 QQQPMWRSSQEKDEKEKEAEKEEKEPEVE-----NDKEDLKEKTDGSDENDEKE 562
Db 2913 LSGSSSSSHADRRHREKREKREKREKREKREKREKREKREKREKREKREKREKREK 2970
QY 563 AVASKRKTANSQGR--KGRITRSMANBANSEAITPOQSAELASMELNNESSRWTEEE-- 619

QY 2038 LRSGLGSSYSPEGVESP-VSSPSLTH-----DKG-----LPKLEELDKGSHLEGE 2085
 Db 5008 LAKPNY---SYAPTVLTPLPVAVQQQQQLYKQNNQKGAQIQMPPH-----GI 5054
 QY 2086 LRPKQFPGVKLGGEAAHL-PHLRPLPESQP--SSSPLL-----QTAPGVKGHRVVTIA 2136
 Db 5055 IMPTEGMLLQKLPALHLPQHQHLPNPPPGKPNPVLHGLQSGQIMPGSVGPPVSA 5114
 QY 2137 -----QHISEVITQDTRHHFQQQLSAPLPAPLYFPFGASCVPILDRPPSDLYLPPDPH 2190
 Db 5115 VLKTAQQQVNSVV-----PVAGIRTAIPNIS-PQSOPRVSP--LVLPFGIS 5157
 QY 2191 GAPARGSPHSGKSPENKTSVL-----GGGEGTEPVSPPEGMTEPGHS--RS 2241
 Db 5158 GVPFDPASLUNDSTYRGVATSRFMLYQHLMKGGDYDDKMGSSPPLLEARRGSGPRRTIA 5217
 QY 2242 YPLLVRDGGTQFSPRMSKSP-----GNTSQPAFFSKLTES---NSAMVSKKKQKQIN 2291
 Db 5218 VP-----HSLQSQDRTAADSQMAQVYVVENTRIPPAHFSEIASRGLYDSGALQLEPPPAH 5273
 QY 2292 KKLNTNRNEPEVNIQSGTEIFNMPAINTGTLMTYRSQAVQEH-----STNWGLEAL 2345
 Db 5274 RPTATISVVVPQ-----QMPAVSSGSPFFIGRDSGVQFGSHHHPGKAMDQMDM 5322
 QY 2346 IRKALMGK--YDQWRESPPLSANAFNPLNASLPAAMPITAADGRSDHLLTSPG 2399
 Db 5323 DRMSIAAVVQQQQLHP-----ALPAGMEL--ASQAPPAMAPP 5363

RESULT 45
 ABP69375
 ID ABP69375 standard; protein; 1708 AA.
 XX AC ABP69375;
 XX DT 20-JAN-2003 (first entry)
 XX DE Human polypeptide SEQ ID NO 1422.
 XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
 KW antiarthritic.
 XX OS Homo sapiens.
 XX PN WO200270539-A2.
 XX PD 12-SEP-2002.
 XX PF 05-MAR-2002; 2002WO-US005095.
 XX PR 05-MAR-2001; 2001US-00799451.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Zhou P, Goodrich RW, Auandi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
 PI Wehrman T, Wang J, Wang D, Drmanac RT;
 XX DR WPI; 2002-759812/82.
 XX DR N-PSDB; AB211592.
 XX PT New polynucleotides comprising sequences assembled from expressed
 PT sequence tags (ESTs), useful for treating cell-proliferative,
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
 PT or coagulation disorders.

PS Claim 9; SEQ ID NO 1422; 1012pp + Sequence Listing; English.
 XX The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences (AB211119-
 CC AB212066) or their mature protein coding portion, active domain coding
 CC protein or complementary sequences. The polynucleotides are useful for
 CC identifying expressed genes or for physical mapping of human genome. The
 CC encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight
 CC markers, as a food supplement, for generating antibodies, in medical
 CC imaging, screening and diagnostic assays and for treating cell-
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
 CC arthritis, etc. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIFO at ftp.wifo.int/pub/published_pct_sequences
 XX Sequence 1708 AA;
 SQ

Query Match 3.5%; Score 458; DB 5; Length 1708;
 Best Local Similarity 20.0%; Pred. No. 1.2e-16;
 Matches 408; Conservative 223; Mismatches 673; Indels 736; Gaps 102;

QY 666 LQOHKLKMEKERNARRKKKAPAAASEAAPPVVEDEMEASGVSGNEEMVEAEALH 725
 Db 1 MMHYKLTGTTLETLPLRKRKEKSLGDEE--PPAFESTKQFSGS-----42
 QY 726 ASGNEVPRGECGPATVNNSDTESIPSPHTBAKDTQONGPKPPATLGADGPPGPPPTP 785
 Db 43 -----PGSDAARNLPLESTKSP-APPSKSV-----PSLEGTGFQRTPKP 83
 QY 786 -----PRTSR--APIETPASEATGA-----PTPPAP-----812
 Db 84 GSGSESGKERRTTSKEISVIQHTSFKSDSLQPSGLEGEDKPLAQFPPPPAPHGRSA 143
 QY 813 -----PPSAPPVVPVKEEKEETAAAPPVEEGEEQKPPAAEALAVDTGAEPEV 862
 Db 144 HSLQPKLVQPNIQVPEILVTEPRDPTTEPPPEKPEK-----TESFQWQ 191
 QY 863 KSECTEE--AEEGPAKGD---AEAEATAEALKAKEKGGSGRATTAKSGAPQDSDS 917
 Db 192 RSOTLAQLPAELPKPKKRLRLAEMAQSSGESSFESSVP---LPRSPQESNVLSGSR 248
 QY 918 SATCSADEVDENAGDGKRLLSRP-----SLTTPGDP-----RANASPKP-----LDL 963
 Db 249 SASFERDDHGKAEAPSPSSDMRPKPLGTHMLTVPSHHPHAREMRKSASEQSPNVSHAHM 308
 QY 964 KOLKORAAAIPPIQVTKVHEPPREDAAATPKAPPAPP-----PPQNLQ-----P 1007
 Db 309 TETRSKSFYDGLSLT-----GFSAPAPVAPPARVAPERRKCFVLVQASLSRPP 358
 QY 1008 ESD---AP--QQGSSPRGKSRSPAPPADKEAFAAQAQLPGDPPCWTSGLPFP-----V 1057
 Db 359 ESELEVAPKGRQSEEPQSSSKSAKSLSSQISSAATSHGGPGGPGGDRPPLGPTV 418
 QY 1058 PPREVKA-----SPHADP-SAFSYA-----PPGH-----1082
 Db 419 FYTEALQVPHFHVQATPLHEKFPVLPVPSLFSFQHLVQHEPGQSPFEFTQAMSLSSP 478
 QY 1083 -----PLPLGLDHTARPVLPRPPTISNP-----PPLISSAKHPSVLBRQI---CA 1124
 Db 479 YSMPLPPLSLFOA--PPLPLQFTVLHPQLHLPQLMHPANIPFRQPSFLPMPYPTSSA 536
 QY 1125 ISQGM-----SVQLHVYSEHAKAPGVPTMGLPLMDPKKLAPFSGVKQE 1170
 Db 537 LSSGFFPLQSQFALQPLGPDVESHLPQIKTSLAPLATGSAG-----LSPSTEYSSD 587
 QY 1171 QLSRPGQAGPESLGVPTAQEASVLRGTALGVSFGSITKGP---STRVPSDSAITVR 1226
 Db 588 IRLPP---VAPPASSAPTSAPP-----LALPACPDWTWVSLVVPVRVQTNMFYSAMYT 639

QY 1227 G-----SITHGTPADVLY-----KGTITRIIGED-----SPSRLDRGREDSLPKG 1266
Db 640 TLSQILVTOGSSATVALPKFEEPPSKG--TTVCAGADVHEVGPSPGSEBOSAPPTP 697
QY 1267 H-----VIYEKKGHVLISYEGGMSVTQCSKEDGRSSGPPHETAAPKRTYDM-----MEGR 1317
Db 698 YLRVPVTLPERKGTLSSESILSL-----EGSSSTAGGSXRVLSPAGSLELTWETOQOKR 752
QY 1318 VGRAISSASTIEGLMGRAPP-----ERHSPHLLKEQHHRGSIITQGPISRYVEA 1366
Db 753 VKEEASAKDEKL--ELVCKSVVLSTEDGKPEKSHL-----GNQOQG-----795
QY 1367 QEDYLREAKLLKREGTTP-----PPPSRDL-----TEAYK--TQALGPLKLPKH 1411
Db 796 -----RELEMLSSLSDDPDTKEIPLPHPALSHGTAPGSEALKYPOPSG-----KPHR 846
QY 1412 EGLVATVKEAGRSIHIEPRELHRTPELP--LAP--RPLKEGSIITQGTPLKYDTGASTTG 1467
Db 847 RGLTPL-----SVKEDSKEDQDPLSLAPSSILPSE-----TSRPAKSEQE---TD 891
QY 1468 SKKHVRSLLIGSPGRTFPVPH-----PLDVM--ADARALERA--CYEESLKS 1510
Db 892 SKK-----VLQFSLHTTTNVSWCYLVNPKPHIQHADRRSSVYAGWCISLYNPN 941
QY 1511 RPTASSGSGSIARGAPVIVPELGKPRQSPLTIEDHCAPAGHLPRGSPVTMRPTRLQ 1570
Db 942 LPGAUSTKAALSLR-----SKQVSKETTYMATAP---HPEAGRLVPSSSRRPRMT 989
QY 1571 E---GSLSSSKAQDKRLTSTPREIAKSPHSTVP---EHHPHDISYVHLLRGVSGVDLY 1624
Db 990 EVHPLSVLSEGGQDLARVKEBERGEPEEPASQORGEPAKIKFE-----GG---Y 1040
QY 1625 RSHIPLAF-----DPTSIPIRGIPDLAAAAYILPHL-----APNPT 1660
Db 1041 KSNEEYVYVGRGKGVVCECGIRCKPSMLKXKHITHDVPYCKCHFAFKTGNL 1100
QY 1661 YPHLYPPYLRIGVPTAALENRQTIINDYITSOHMHTATATAQADMLRGLSPRESSL 1720
Db 1101 TKHMKSAHKSCKQETGVLE-----ELEAEETS-----DOLFQDSEGREGE 1143
QY 1721 ALNYAAGPRGIIDLSQVPHLPVLVPTPGTATAMDRLAYLPTAPQPFSSRHSSSLSPG 1780
Db 1144 AV-----BEHQFSDL-----1153
QY 1781 GPHTLTKPTTSSSERDRDRDREREKSIITSTTT-----VEHAPIWR 1828
Db 1154 -----EDSDSDSLDEDEDEEESQDELSPSEAPPGPHPALRADSSPILG 1202
QY 1829 P-----GTQSSGSSGS-----SGGGGSSSRPA-----SHSHAHQ 1859
Db 1203 POPPDAPASGTEATRGSSVSEABRLTASSCSMSQMPGLPWLGPAPLAGSVEKDTGSALS 1262
QY 1860 HSPISPR-----TQDALQORPSVLNHTGMKGIITAVEPSKPTVLRSTSTSPV-- 1907
Db 1263 YKPVSPRRPMSPKESGRPLARKHSLTKNDS-----SPORCSPAKEPOASAPSP 1314
QY 1908 -----RPAATFPFATHCPILG-----GTLDCGVYPTLMEPVL 1937
Db 1315 GLHVDPCRGMGPLPCGSPRLQLSPLTCLGLRELAPRAHVLSKLEGTTD--PG-----1365
QY 1938 LPKEAP-----RVARPERPRADTGHAFLAKPPPARSGLEPASPSPKGSB-----PRP 1983
Db 1366 LPRYSPTRRSPQAESPPR-----SAPPKGWALAGPSGAGEHGPGLGLAPRV 1415
QY 1984 LVPPVS--GHATARTAKNLAHPHA--SPDPPAPASADPHREKTOSKPSIQEELRSL 2041
Db 1416 LFPFAPLPHLLRSRSPETCASPKWASRSRSCSPGFA-----HPLSRPFS-----ALHDF 1466
QY 2042 GYHGSSVSPGVSPVSPVSPSLTHDKGLPKHLEELDKSHLEGELRKPQGP--VKLGGEA 2100
Db 1467 HGHLALTEINI-----FSH---LPLHSQLTRA-----PCPLIPICG-- 1501
QY 2101 AHLPHLRPLPESQSSSPLLOTAP-----GVKGHQRVVTVAQHSIVITQDYTRHHPOQ 2154

Db 1502 -----IQVQARPGAHPHTLLPCPTAAWVSGSGGSDLTGARE-----AQEGRWSPT 1550
QY 2155 LSAPLPAPLYSPFGASCPVLDLRRPPSDLYLPPDPH-----GAPARGSPHSEGG 2203
Db 1551 SSS-----ASVSPVAKVSKFTLSSELEGDYPKERERTGGGPPDPWTPHGTGA 1600
QY 2204 KRSPEPKTSLVGGSDGIEPVPSPGEMTEPHGHSRAVYLLYRDGEQTEPSRMGSKSPG 2263
Db 1601 PAEPTPTH-----PCTPPDTLPRPPOGR-----RAAQSWSPRLSPRAPA 1641
QY 2264 N---TSQPAFPSK-----LTESNSAMVSKKQINKKLANTH-----NRNEPEYNISOP 2309
Db 1642 NPESATPPDLDRSSVGCGLAEA--SARFPARTNLNLSGEPRTRQDSPKPSGSGEPRAHPHQ 1700
RESULT 46
AAB41231
ID AAB41231 standard; protein; 2971 AA.
XX AAB41231;
AC AC
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF995 polypeptide sequence SEQ ID NO:1990.
DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000NO-US008621.
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
DR N-PSDB; AAC75440.
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX Claim 11; Page 1510-1517; 5507pp; English.
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;

dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an OREX-associated disorder. The nucleic acids can be used to express OREX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 2971 AA;

Query Match 3.5%; Score 456; DB 3; Length 2971;
Best Local Similarity 19.6%; Pred. No. 3.3e-16;
Matches 596; Conservative 305; Mismatches 1141; Indels 1006; Gaps 134;

QY 71 EFQPGNERSQE---LHLRPESHSLPELKGEMEFI--ESKRPLELLPD--PLLRSP 122
DB 91 DFQPGDEDEDEETIEVEEQEGNDAAQREIEULRREGELPEELRSLPPQLLEGP 150
QY 123 LLAATGQAGSEDLTKDRSLTGKLEVPSPSPHTDPELELVPPR---LSKEELIQMDRV 179
DB 151 SSPSQTPSHSDTRDPEGAE-----EPQV---LEIKPPSAVTQKNQPHDPD 202
QY 180 DREITWVQOISKLK---KKQOLEEBAKPEPEKVPSPPIESKHSVLQIYY----- 231
DB 203 DEEFTANEBAEEDETIAAEQLEGEVDHAMELSLAREGEL-----SMEELQOYAGA 257
QY 232 -----DENRKKAAHRILEGGP-QVELPLYNQPSDTRQYHENIKINQAMKKLI 281
DB 258 YAPGSSSEDEDEVDANSSDCEPGPVEAEEP-----POEDSSSQSDSVE----- 304
QY 282 LYFKRRNHARKQKFCORYDQLEALBKVKVERIENNRRAKESKRYVEYKQPPETR 341
DB 305 -----DRSEDEDEHEEETSGSSASESESESEDAQSQSOA 343
QY 342 KORE-----LOERMQRVQGGSLGMSAARSEHEVSEIIDLSEQENLEKQMRQ 391
DB 344 DEEEEDDFGVEYLLARDEBQSEADAGSPPTPGTTLGPKKEITDIAAAESLOPKGT 403
QY 392 LAVIP-----PMLYDADQRIKEI-----NMGIMADPMKYKQROVMNN----- 431
DB 404 LATTQVKTPILLRGLRQREYQHIGDLVYMEKKLNGILADEMGLGKTIQTIISLAHL 463
QY 432 -----WS-----BOKEETFREKFMQHPKQFGLIASFLERKTVAE----- 465
DB 464 ACEKGNWGPILLIIVPTSVMLNMEMELKWCPSF---KILTYGAQKERKLRQGWTKEN 519
QY 466 -----CVLYYLT-----KKNENKSLVRRSYRRRGKSQOQOQOQOQOQOQOQO 507
DB 520 AFHVCTISYKLVLDHQAFRRKNRYLIILDEAQINKFKSQWQSLNFNFSORLLLTGT 579
QY 508 -----QQQPMRSSQEEKDEK-----KEKAKEKEE-----KP- 536
DB 580 PLQNSLMELWSLMFLMHPVQSHREFKWFNSNPLTGMIEGSOEYNEGLVKRLHKVLRFP 639
QY 537 ---EVENDEKEDLLKEKTD-----DTSGEDNDEKEAVASKRKTANSQGRR 578
DB 640 LLRRVKVDVEKQPKKYEVHRCLSQRQRCCLYDDFMAQTTTKETLAT-----GHF 690
QY 579 KGRIT-----RSMANBAN-----SEEAITPOQSAELASMEINSSRWTEEMETAKG 626
DB 691 MSVINILMQLRKVCNHNPLFDPRVTPSPFITPGICFSTASLVLRATDVHPLQIDMGRED 750
QY 627 LLE-HGRNWSAIARMVSKTVSCKNFYNYKKRQNLDELQOHLKMKERNAREKKK 685
DB 751 LIGLEGR-----VSRYEADTFLPRHL-----SRRVLLE 779

686 APAASEEAAAPPVVEDEMEASGVSGNEEMVEEAEALHASGNVPRGECGSPATVNS 745
DB 780 VATADPP-----PPRPKPVKMKVN-----RMLQ-----EVPKQEGRTVVVWNP 818
QY 746 ---SDTESISPHTE-AAKDTGQNGPKP---PATLGADGPPGPP---TPRRTSRA 792
DB 819 RAPLGFVVRPPPGPELSAQPT--PGPVQVLPAASLMVASPAGPPLIPASPPGVLPLP 876
QY 793 PIEPTPASEATGAPTP-----PPAP-----PSPAPPVVPKKEEETAAPVVE 838
DB 877 PLOPNSGSLPQVLPSPGLVLSGTSRPPPTLSLKTTPAPVRLSP-----APPG 926
QY 839 EGEEQKPAABELAVDTGKAEPVKSECTEAEPEGAPAKODAAEATAAGALKAEKKG 898
DB 927 PSSLLKP-----LTVPPGYTFPP----- 944
QY 899 GSGRATTAKSGAPODSDSSATCSADEVDEAEGGDKNRLSPRPSILLTPTGD----- 950
DB 945 ---AAATTT-----STTATATTATVAPTPAPQRLILSPDMQARLPSEGVVSIQILA 994
QY 951 -----PRANASPKQLDLK-----QLKQ-----RAAAIPPIQVTK-----V 981
DB 995 SLAQRPVANAGSKLTFQIQGNKLTLTCAQVQLAVGQPRPLQMPPTMVNTGVVIV 1054
QY 982 HEPPREDAAATKPAAPPAPPPQNLQSPESAPO---OPGSSPR---GKSRSPAP----- 1028
DB 1055 ROAPRDGLTPVPLAPAPRPPSSGLPAVLNRPRTLTPGRLTPTLTGTARAPMTPTLVRP 1114
QY 1029 -----PADKEAFAAEAQKLPDPCWTSGLPFPVP-----PREVIKASHAPDP 1072
DB 1115 LKLVHSPSPVESASA-----PGAAPL-TISSPLHVPSLPGPASPMPIPNSSPLASPV 1168
QY 1073 SAFSVAPPGLPLGLHDTARVLPRPPTISNPPPLISSAKHPSVLERQIGAISQMSVQ 1132
DB 1169 STVSVPLSSSLPIGVPITLPAASAPLTIPISAPLTVSASGALL----- 1214
QY 1133 LHVPSYSEHAKAVGVVTWGLPLPMDPKKLAPSGVKQEQOLSPPRQAGPPESLGVTQAQ-- 1190
DB 1215 -----TSVTPLPAV---VPAAPGPPSLAP--SGA-----SPSASA---LTIGLATAPSL 1256
QY 1191 EASVLGRTALGVPGGSIKGIPTSRVPSDSAITYRGSITHG-----TP 1234
DB 1257 SSSQTFGHPHLLAPTSSSHVPLGNSLVAFACSPVLVPASALASFPFSAFNPAPQAASLLAP 1316
QY 1235 ADVLKGTITRIIGBDSRSLDRGREDLSLPGKHVIY---EGKKGHVLSYEGMSVYQCS 1290
DB 1317 ASSASQALATPLAMAAQPTAILAPSPAPPLAPLVLAPSGAAPVLASSQTPVPWAPS 1376
QY 1291 KEDGRS---SSGP---PHETAAPKRYDMMEGRVGRAL--SSASIEGL-MGRAPTPERHSPH 1343
DB 1377 STFGTSLASASVPVAPTPVLAPSSITQMLPAPVPSPLPSPASTQTLALAPALAPT----- 1431
QY 1344 HLKEQHIRGSIQIGIPRSYVEAQEDYLRRKALKREGTTPPPPSRDLTEAYKTQALG 1403
DB 1432 -----JGGS-----SPSQTLSLGTGNPOGPPPTQTL----- 1458
QY 1404 PLKLPAPHEGLVATVKEAGRSIHEIPRELRHTPELPLAPR-PLKEGSIQGTPLKYDTG 1462
DB 1459 ---LTPA-SSLNPT---PAQTLSLAPGPPGLGTQTLAPAPLAPASVPGPAPAHHTLT 1511
QY 1463 ASTTQSKK---HDVRSLIGSGPRTFPVPHLDVWADAPALERACRYEESLKRPGTASS 1518
DB 1512 APASSASILLAPASVQTLTSPA-----PVTTLGPAAAQTLALAPASTQSPASQASL 1564
QY 1519 GCSIARGAVIV-----PELKGPROSLTYEDHGAFFAGHLPGRGSPVMTREPT 1566
DB 1565 VVSAGAAPLFTVMVSRVLVSKDEPDTLTLRSGPPSPPTSTATSGGPRRROP-----PP 1620
QY 1567 PR-----LOE-----GSLSSSSASQDRKLTSTFRIAK---SPH 1597
DB 1621 PRSPFVLDLSEKKRQRSERLERIFOLSEAHGALAPVYGVTEVLDFCTILPQPVASPIGR 1680
QY 1598 STVPEH-----HPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSIPIGILPDA 1647

RESULT 48
ID ABB59410
XX ABB59410 standard; protein; 1612 AA.
AC ABB59410;
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 5022.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
DR N-PSDB; ABL03513.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 5022; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1612 AA;
SQ
Query Match 3.4%; Score 452; DB 4; Length 1612;
Best Local Similarity 21.8%; Pred. No. 2.5e-16;
Matches 419; Conservative 149; Mismatches 748; Indels 604; Gaps 90;
QY 560 EKEAVASKGRKTA-----NSQRRKGRITRSMANEAN---SERAITPQOSAEALASMELE 611
DB 4 ESRAATRGELASASQCNQRRHRER-DRS-ANNINGNAATAAATAGGVHRCGAICE 61
QY 612 ---SSRWTE-----EEMETAKKGLLEHGRNWSAIRMVGSKTVCCKNFYFNKKR 659
DB 62 RWQNRWSHHPLTALVASHQAHAAVGLKCGGKS-----AVCVGLSVQC----- 107
QY 660 QNLEILQOHLKQKERNAR-----RKKKAPAAASEAA-----FPPVVEDE 703
DB 108 --VDDLTTGPVPC--ALNTRCVTDGLEICVPKSAEAAPTAAVAKWVTITDSPPVVSSEA 163
QY 704 EMEASGVSGNEEMVEAEALHA--SGNEVPRCEGSPA-----TVNNSDTSIPS 753
DB 164 PLVDSTNSGSGVEISTEGAPASAVPTKPSIPVETTOAPVETTOIPLETQVAVETTOIPL 223
QY 754 PHTAAKD-----TQONGP-KPPATLGADGPP-----PG-----PPTP-- 785

DB 224 ETTOASGETTTAAIBETTITGSEAPLEPESTVPSDDTTPVDSTLAPGWEGTYPTIEPNTPAE 283
QY 786 ---PRRTSRADIEPTPASEATCAPTPPPAPSPSPAPPPVVPVKEEKEETATAAPPVEEGE 841
DB 284 DNAPAGTTLAPGEVDPS-----PIDNSPLDPNAP-----EESTNEPGLVDPLPAD 332
QY 842 EQKPPAAELAVDTGKAEFPVKSECTEEAEEGPAKGDAAEAATAEALKAKEKGGSG 901
DB 333 TTTAP---DSPVEGSSAAPTADVTTAAPGAPADGSSA-APGSPADGSSAAP---GSP 384
QY 902 RATTAKSSCAPODSSSAT-CSADEVDEAEGDKNRLSPRPSLLTPTGDP--ANASPO 958
DB 385 ADVTTAAGAPADGSSAAPGAPADGSSAAPG-----SPADVTTAAGAPADGSSAAPG 437
QY 959 KPLDLKQLKQRAAAIPIQVTKVHEPPREDAAPTKPAP-----APPPONL-----Q 1006
DB 438 SPAE-----GSSAAGAPADVTTAAGAPADGSSAAPGAPADGSSAAPGSPADVTTAAPGA 493
QY 1007 PESDAPQOPGSSPRGKSRSPAPPADKEAFAEAQKLPDGPCCWTSGLP-----FPV 1057
DB 494 PADGSSAAPGSPAGSSAAPGAPADVTTA-----PGAPADGSSAAPGAPADGSSAAPG 547
QY 1058 PPREVIKASPHAP-----DPSAFSYAPPCHPLPLGLHDTARVLPVLPRTTISNP--P 1106
DB 548 SPADVTTAAGAPADGSSAAPGAPADGSSAAPGSPADV---TTAAGAPADGSSAAPGAP 604
QY 1107 PLISSAKHPSVLERQIGAISQGMVSQVLPVHYSEHAKAPVGP--VTMGLP-LPMDPKKLAP 1163
DB 605 ADVTTAAP-----GAPADGSSAAPGSPAGSSAAPGAPADVTTAAGAPADGSSAAP 656
QY 1164 FSGVKOEQLSPRGQGPPEPSLGVPTAQEAASVLRLGTALG-----SVFGGSITTKIPS 1214
DB 657 --GAPADGSS-----AAPGSPADVTTAAGAPADGSSAAPGSPADVTTAAGAPADGSSPA 710
QY 1215 TRVPDSAITYRGSYTHGTTPADVLYKGTITRIIGEDSPSLRDLGRDLSLPKHVIEGKK 1274
DB 711 PGAPADGS-----SAAPGSPADV----- 728
QY 1275 GHVLSYEGGMSVTQCSKEDGRSS--GPHE--TAAPKRTYDMWGRVGRAISSASIEGLM 1331
DB 729 -----TTAAGAPADGSSAAPGSPAGSSAAPGAPAD-----VTTAAGAPADG-- 772
QY 1332 GRAIPPERHSPHLKEQHIRGSITQIGIPRSYVEAQEDYLREAKLLKREGTPPPPPSR 1391
DB 773 -----SSAAGAP-----ADGSSAAPGSPA 792
QY 1392 DUTEAYKTOALG-PLKLKPAHEGLVATVKEAGRSIHEIPRELRHTPELPLAPRPLKEGS 1450
DB 793 DVT---TAAGAPADGSSAAPGAPADGSSA-----PGSPADADGSS 831
QY 1451 ITQGTPLKYDT---GASTTGSKKHDSRLJGSP--GRTEFPVHPLDVMADARALERACYE 1505
DB 832 AAPGSPADVTTAAGAPADGSS-----AAPGSPAGSSAAPGSPADVTTAAGAP-----A 882
QY 1506 ESLKSRGPTASSGSGSIARGAPVIVPELKGPRQSPLTYEDHGAPFAG-HLPRGSPVTMRE 1564
DB 883 DGSSAAGP-APADGSSAAPGSPADVTTAAGAPADGSSAAPGAPADGSSAAPGSPADVT 941
QY 1565 PTFRLQEGSLSSKASQDRKLTSTPREIAKSPHSTVPEHHPHPIPSYEHLLRGVSGVDLY 1624
DB 942 AAPGAPADGSSAAPGSPADVTTAAGAPADGSSA-----PGSPAEG----- 983
QY 1625 RSHIPLA-FDPTSI PRGIPLDAAAYVLPRLHAPNPTYPHLYPPYLRGVDPDTAALNRQ 1683
DB 984 SSAAGAPADVTTAAGAPADGSSA-----AP-----GAP----- 1013
QY 1684 TIINDYITSQOMHNTATATAOARADMLRGLSPRESSLALNYAAGPRGIIIDLQVPHLPVL 1743
DB 1014 -----AD---GSSAAGSPADVTTAAGAPADGSSA----- 1041
QY 1744 VPPTGCTATAMDRLAYLFTAPQPPSSRHSRSP-----LSPGCPHLTKPTTTSSSERBR 1798

QY 746 -----SDTESIPSHTE-AAKDTQONGKP-----PATLGADGPPPGPP-----TPPRTSRA 792
Db 819 RAPLGPVVRPPPGPELSAQT--PGVPQVLPASLMVMSASPAGPLLPASRRPPGVLLP 876
QY 793 PIBPTPASEATGAPT-----PPAP-----PSPSAPPPVVPKXEKEEETAAAPVVE 838
Db 877 PLQFNGSGLPQVLPPLGVLGTSRPTPTLSLKPTPPAVRUSP-----APPG 926
QY 839 EGBEQKPPAAEELAVDTGKAEPEVKSECTBEABEGPAKGDABEABATAGALKAEKKEG 898
Db 927 SSSLLKP-----LTVPGGYTFP-----944
QY 899 GSGRATKXSGAPQSDSATSADBEAEGDKNRLLSPPSLLTPTGD-----950
Db 945 --AAATTT-----STTTATTTAVAPTAPQRLILSPDMQARLPSGVVSIQOLA 994
QY 951 -----PRANASPOKPLDK-----OLAQ-----RAAIIPIQVTK-----V 981
Db 995 SLAORPVANAGGSKPLTFIQGNKULITGQVROLAVGQRPLOMPTTWNNTGVVKIV 1054
QY 982 HEPREDAAPTKAPPAPPPQNLQPESDAQ-----OPGSSPR-----GKRSRSPAP-----1028
Db 1055 RQAPRDLPTVPLAPAPRPPSSGLPAVLNPRTLTFLGRLPTLTGTAARAPMTPTLVRP 1114
QY 1029 -----PADKEAFABEAOKLPGDPCWTSGLPFPVP-----PREVIKASHPADP 1072
Db 1115 LKLHVSPSPVEBSA-----PGAAPL--TISSLHVPSSLPGPASSPMPPTPNSPLASPV 1168
QY 1073 SAFSYPAGHPPLGLLHDTARVLPPTTISNPPPLISSAKHPSVLERQIGALSQMSVQ 1132
Db 1169 SSTVSPLSSSLPISVPTLTPAPASAPLTIPIISAPLTVSASGALL-----1214
QY 1133 LHVPSYEHAKAPGVPTMGDPLPMDPKLAPFGVQKQESLPRGOAGPPES---LGVPTA 1189
Db 1215 -----TSVTPPLAVVPAAPG-----PSLOPSG-ASPASALTLGLATA 1253
QY 1190 Q---EASVLRGTLGSPVGGSIKIGIPSTRVSPDSAITYRGSITHG-----1232
Db 1254 PSLSSQTPCHPLLAPTSHVPLNSTVAPACSPVLVPASALASPPSPAPNAPAAQASL 1313
QY 1233 -TPADVLYKGTIRIIGEOSPSRLDREGDSLPKGHIY-----EGKKGHVLSVEGMSVT 1287
Db 1314 LAPASSAQALATPLAPMAQPTAILAPSPAPPLAPLVLAPSPGAAPVLASSQTPVPVM 1373
QY 1288 QCSKEGDS--SSGP---PHETAAPKRTYDMGRVGRAI--SSASIEGL-MGRAIPPERH 1340
Db 1374 APSSTGTSLASAPVPAFTPLVAPSTQTMPLPAPVPSPLPSPASTQTLALAPALAPT-- 1431
QY 1341 SPHLKEQHHRIGSITOGIPRSYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQ 1400
Db 1432 -----LGGS-----SPSQTLISLGTGNPQGFPTQTLS-----1458
QY 1401 ALGPLKLKPAHEGLVATVKEAGRSIHEIPREELRHTPELAPR-PLKSGSITQGTPLKY 1459
Db 1459 -----LTPA--SSLVPT---PAQTLSLAPGPPGLQPTQLSLAPAPPLAPASVGPAPAH 1508
QY 1460 DTCASTTGSKK---HDVRSLIGSPKRTPPVPHPLVDADARALERACYEESLKSPPGTA 1515
Db 1509 LTLAPASSASLAPASVQTLTUSPA-----PVTILGPAQAOTLALAPASTQSPASQA 1561
QY 1516 SSSGGSIGARGAPVIV-----PELGKPRQSPLTVEDHGAPFAGHLPRGSPVTMR 1563
Db 1562 SSLVVSASGAAPLVMTVMSRLPVSKDEPDTLTLRSGLPPSPSPSTATSFSGPRPRRQP--- 1617
QY 1564 EPTPR-----LOE-----GSLSSSKASQDKLTSTPRETAK---1594
Db 1618 PPPPRPFYLDLSLEEKRRQRORSLERIRIFOLSEAHGALAPVYGTVELDFCTLPPVASPI 1677
QY 1595 SPHSTVPEH-----HPHIPISVPEHLRGVSGVDLYRSHIPLAFDPTSTIPRGILD 1644
Db 1678 GKSPGSPHPTFTTYTEAAHRAVLFPQORLDQJSEIERIFVMP-----PVBAPP 1729
QY 1645 AAAAYLPRHAP-----NPTYPHLYPPYLI-----RGYPD-----TA 1677

Db 1730 SLHACHPPWLPAPROAAFOELASELWPRARPLHRIVCNMRTQFPDLRLIYDCGLQTL 1789
QY 1678 ALENRO-----TIINDYITSQ-----QMHNTATAMARADMLRG---1712
Db 1790 AVLLROLKAEHRVLIFTQMTMLDVLVEQFLYHGHLYLRLDOSTRVEQOALMERFNAD 1849
QY 1713 -----LSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPTPTGTATAMDR-----1757
Db 1850 KRIFCFILSTRSGGVGNLTGADTVVFYDSW-----NPTMDAQADRCHRIGQTR 1900
QY 1758 -----LAYLPTA-----POPFSSRHS 1773
Db 1901 DVHIVELISERTVEENILKANOKRMLGDMAEJGNGFTTAYFKQQTIRELFDMPLEEPSS 1960
QY 1774 SS-PLSP-----GGPHLTKTPTTSSSERDR-----1800
Db 1961 SSVPSAPEEBETVASKQTHILEQALCRAEDDEDIRAATOAKAEQVAELAEFNENDGPPA 2020
QY 1801 -----DRERDRDREREKSILTSTTVEH-----1823
Db 2021 GEGERAGRPGAEDDEMSRAEQEIAALVEQLTPIERYAMKFLEASLEEVSRUELQAEQV 2080
QY 1824 -----APIWR--PGTEOSSSG--SSCGGGSSSRPASHASHAQH---SPISPR 1867
Db 2081 EAARKOLDQAKBEVFRLPQEBEGPGAGDESSCGTGGTHRRSKAKAPERPGTRVSERL 2140
QY 1868 QDALQORPSVLNHTGMKLIITAVEPSKPTVLASTST---SSPVRPAATPPATHCPLOGT 1924
Db 2141 RGAETQGANHTP---VISAHQ-----TRSTTTTPRCSPARERVPAPRPRTPAS 2190
QY 1925 LDGVPTLMEPVLLPKAEAPRVARPERPRADTGHAFKAPKPARSGLEPASPSKSGSPRPL 1984
Db 2191 APAAPAL---VPVPSAPVPIASAPNITILPVHILPSPPPPSQIIPCCSPA--CTPPPA 2245
QY 1985 VPPVSGHATIAITPAKNLAPHASP---DPPAPPASAS-----DPHREKTOQSFPSI 2033
Db 2246 CTPPPAHTP---PPAQCTLVTPSSPLLGLPPSPVPIASVNTLPLGLRPAELCAQALASP 2302
QY 2034 QELERSLGVHSS---YSPGVREVS---PV-----SSPSLTHDKG-LPKHLE 2075
Db 2303 ESLELASVASSSETSSLSLVPKDLLPVAVEILLPVSEKNLSLTPSAFSLTLEAGSIPNGQE 2362
QY 2076 ELDKSHLEGELRPQKPGVYKLGAEAAHP-----HLRPLPESQPSSSPLLQTPAGVKGH 2129
Db 2363 QEAPDSAECTTLTVLP-----EGEELPLCVSENGLELPPSAASDEPLQBLEADR-- 2413
QY 2130 QRVVTLAQHISEVITODYTRHHPOQL-SAPLPAPLYSPFGASCPLVDLRRPPSDLYLPPP 2188
Db 2414 -----TSEBELTEAKTPTSSPEKPELVTAEVAAPSTSSATSSP-----2452
QY 2189 DHGAPARGSPHSEGGKRSPEPNKTSVLGGEGDGBVPSPEGWTEPFGHSRAVYLLYRD 2248
Db 2453 -----EGPSPARPPR-----RRT 2465
QY 2249 GEQTEPSRMGSKSPGNTSQPPA--PFSKLTENSAMVKSKQKOEINKLANTHNRNEPEYNI 2306
Db 2466 SADVEIRGQGTGRPG---QPPGPKVLKPLGRLVTVVEEKELVQRRR-----2509
QY 2307 SQFGTEIFNMPIAITGTGLMYRSQAOVEHASTNMGLEAIIIRKALMKYDQWESPPLS-- 2364
Db 2510 QQRGAASLTVPGVSET-----SASPGSPSV--RSMSPG---ESSPPIGGP 2549
QY 2365 -----ANAFVPLNASLPAAMP---ITAADGRSDHTLT 2395
Db 2550 CBAAPSSSLTPPQOPFIARRHIELGTGGGPGENGDGALLAITPPAVXRRRRPPKKNR 2609
QY 2396 SPGGGKAKVSGRPSRRKAKS---PAPG-----LASGDRPPSVSVHSEG 2437
Db 2610 SPADAGRVDEAPSTLCKTKTGADVPQFETLIVADPVLQELIPGQPLGPQPH--- 2666
QY 2438 DCNRRTPLTNRVWE---DRPSSAGSTPPFPYPLIMLRLOAGVMA-----PPPPGLP 2485

Db 2667 ---RNPPLSPVKKRRGRPPKARDLP.PIP-----GTISSAGDGNSESRTOPPHP 2713

QY 2486 AGSGPLAGPHHAWDEPKPLLC 2507

Db 2714 SPLTPL-----PPLVLC 2725

RESULT 50

AAB50363

ID AAB50363 standard; protein; 2972 AA.

AC AAB50363;

XX

DT 12-MAR-2001 (first entry)

DE Human SRCAP.

XX

XX Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;

KW CAMP regulatory element; CREB binding protein; CBP; ATPase;

KW transcription activation; DEAD box RNA dependent helicase;

KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.

XX

OS Homo sapiens.

XX

XX WO200073467-A1.

PN

XX

XX 07-DEC-2000.

XX

XX 25-MAY-2000; 2000WO-US014719.

XX

XX 27-MAY-1999; 99US-0136620P.

PR

XX 25-MAY-2000; 2000US-00579181.

XX

XX (UYSL-) UNIV SAINT LOUIS.

XX

XX Chirivia J, Yaciuk P;

PI

XX

XX WPI; 2001-061545/07.

DR

XX N-PSDB; AAC89860.

XX

XX

PT Snf2 related CAMP regulatory element (CREB) binding protein (CBP)

PT activator protein, capable of co-activating CREB binding protein, useful

PT for modulating transcription and for affecting viral infection.

XX

PS Claim 1; Page 86-94; 103pp; English.

XX

XX The present sequence is an Snf2 related CREB (cAMP regulatory element)

CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has

CC ATPase activity and is capable of activating transcription. SRCAP

CC polypeptides are useful for activating transcription in a cell, for

CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated

CC activation of transcription in a cell, for treating a patient having a

CC disease involving a function such as insufficient transcription of a

CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent

CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor

CC affected by SRCAP protein. Compounds that modulate SRCAP function, such

CC as antibodies, antisense molecules, polynucleotides or ribozymes, are

CC useful for treating diseases mediated by SRCAP-activated transcription,

CC for example, infection by adenovirus, hepatitis C virus, human

CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or

CC hepatitis B virus

XX

XX Sequence 2972 AA;

SQ

Query Match 3.4%; Score 452; DB 4; Length 2972;

Best Local Similarity 19.0%; Pred. No. 5.5e-16;

Matches 587; Conservative 296; Mismatches 1107; Indels 1092; Gaps 130;

QY 71 EFQPGNRSOS---LHLRPESHSLYLPGLKSEMEFI--ESKRPRLLELPD--PLLRPSP 122

Db 91 DFQOEDEDEDEETIEVEQEGQNDAAORREIELLRREGELPLELLRSLPQLLEGP 150

QY 123 LLATGQPAGSEDLTKRSLTKLEPVPSPSPPHPTDPELELVPPR---LSKEELIQNMDRV 179

Db 151 SSPSQTPSSHSDTRDPEGAEE-----EPQV---LEIKPPPSAVTQRNKPQWHPDED 202

QY 180 DREITWVQOISKLK---KKOOLLEEAAKPEPEKPVSPPPPIESKHSLVLIY-----231

Db 203 DEEFTANEAEAEDEEDTIAAEQEULEGEVDHAMELSAREGEL-----SMBELLOQYAGA 257

QY 232 -----DENRKKAEEAHRILEGLP-QVELPLYNQPSDTRQVHENIKINQAMRKKLI 281

Db 258 YAPGSGSEDEDEDEVDANDSDCEPGEVAAEP---PQEDSSQSDSVE-----304

QY 282 LYFKERNHARKQKQFCORYDQMLEALEKKVERIENPNRRRAKESKVREYVEKQFPEIR 341

Db 305 -----DRSEDEDEHESEETSGSASESESESESEDAQSOSA 343

QY 342 KORE-----LOERMQSRVQSGSLNSAARSEHVESEIIDGISEQENLSEKQ 391

Db 344 DEEEDDDDFGVEYLLARDEEQEADAGSGPTTGGTIGPKKKEITDIAAAESLPKGYT 403

QY 392 LAVIP-----PMLYDADOORIKFI-----NMNGLMADPMKVYKDRQVMNM---431

Db 404 LATQVKTPIPLLRLGQRLREYQHIGLDWLVWYKELKNGILADENGLKTIQTISLAHL 463

QY 432 -----WS-----EOKETFRKFMQHPKPNFGLIASFLERKTVAE-----465

Db 464 ACEKGNWGPHLIIVPTSVMLNWMELKRWCFSP---KILTYGAQKERKLRQGWTRPN 519

QY 466 -----CVLYYYLT-----KKNENYKSLVRRSRYRRGKSGOQQOQQOQQO-----507

Db 520 AFHVCIITSYKLVLODHQAFRRKNWRYLILDEAQNKNFKSQRWQSLMNFNSORLLTGT 579

QY 508 -----QQQPMRPSQEEKDEKE-----KEKEAEKEEE-----KP-536

Db 580 PLQNSLMELMSLMHFLMPHVFSHREFKEWFSNPLTGMIEGSGQYNEGLVKRLHVLRF 639

QY 537 -----EVENDEKEDLLKEKD-----DTSGEDNDEKEAVASKRGTANSQGR 578

Db 640 LLRRVKVDVEKQMPKKYEHVIRCLSKRQRCLYDDFMATTTKETLAT-----GHF 690

QY 579 KGRIT-----RSMANEAN-----SEEAITPQSAELASLMELNESSRWTEEMETAKG 626

Db 691 MGVINILMQLRKVCNHPNLFDPRPVTSPTTIGICFSTASIVLRATDVHPQLQIDMRFD 750

QY 627 LLE-HGRNWSAIAARMVGSKTVSQCKNFYNYKKRQNLDEILOOHKLKMEKERNARKKK 685

Db 751 LIGLEGR-----VSRYEADTFLPRHL-----SRRVLE 779

QY 686 APAASEEAAFPVVEDEMEASGVSGNEEEMVEAEALHASGNEVPRGECSPATVNS 745

Db 780 VATAPDP-----PPRPKPVKMKVN-----RMLQ-----PVPKQEGRTVVVVNNP 818

QY 746 -----SDTESIPSPHTE-AAKDTGQNGPKP---PATLGADGPPPGPP-----TPPRTSRA 792

Db 819 RAPLGPVVRPPPGPGLSAOPT--PGPVQVLPASLWSASAGPPLIPASRPPGPVLLP 876

QY 793 PIEPTPASEATGAPT-----PPAP-----PSPSAPPVVPVKEEKEETAAPPYE 838

Db 877 PLQPSGSLPQVLPSPGLSVLGSRTSRPTTSLKPTPPAPVRLSP-----APPG 926

QY 839 EBEQKPPAAEBELAVDTGAAEPVSEKTEEAEEGPAKGDAAEABATAGALKAEKKG 898

Db 927 SSSLLKP-----LTVPPPGYTFPP-----944

QY 899 GSGRATTAKSGAPODSSSATCSADEVDEAGGDKNRLSPRPSLLTPTGD-----950

Db 945 ---AAATTT-----STTATATTAVPAPPAQORLILSPDMQARLPSGEVVSIGOLA 994

QY 951 -----PRANASQKPLDLK-----OLKQ-----RAAIPPIQVTK-----V 981

Db 995 SLAQRPVANAGSKPLTFQIQGNKLTLTGAQVRLAVGQFRPLQMPPTVMVNTGVVIVV 1054

QY 982 HEPREDAAPTAPAPAPPPQNLQESDAPQ---OPGSSPR---GKSRSPAP-----1028

Db 1055 RQAPRDLGTPVPLAPAPRPPSSGLPAVLNRPRTLTCRLPTTLTGARAPMTPTTLVRP 1114
QY 1029 -----PADKEAFAAEAKLPGDPPCWTSLGPPVP-----PREVIKASHPADP 1072
Db 1115 LKLVHSPSPSEVSASA-----PGAAPL-TISSPLHVFPSSLPGPASSPMPIPNSPLASPV 1168
QY 1073 SPSYAPPCHPLPLGLHDTARVPLPRPPTISNPPLISSAKHPSVLRQIGALISQGMVQ 1132
Db 1169 SSTVSPLSSSLPISVPTTLPAAPASAPLTIPIASPLTVSASGALL----- 1214
QY 1133 LHVPSYSEHAKAPVCPVTMTGLPMDPKLAPFGSVQKQLSPRGQAGPPES-----LGVPTA 1189
Db 1215 -----TSVTPPLAPVPAAPG-----PSLPQSG-ASPSASALTGLGATA 1253
QY 1190 Q-----EASVLRGTALSGVPGSITKIGIPSTRVPSDAITYRGSITHG----- 1232
Db 1254 PSLSSSTQPGHPLLLAFTSSHVPLNSTVAPACSPVLVPASALASPPSPAPNAPQAASL 1313
QY 1233 -TPADVLYKGTIRIIGEDSPSLDRGREDLSLPGHVIY-----EGKGHVLSYEGGMSVT 1287
Db 1314 LAPASSASQALATPLAPMAAPQTAILAPSPAPPLAPLPLVLAPSPGAAPVLASSQTVPVVM 1373
QY 1288 QCSKEDGRS--SSGP---PHETAAPKRTYDMMEGRVGRAI--SSASIEGL-MGRAIPPERH 1340
Db 1374 APSTPGTSLASAPVPAPTPVLAPSTQMLPAPVSPPLSPASSTQTLALAPALAPT-- 1431
QY 1341 SPHLKEQHHRISITOGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTQ 1400
Db 1432 -----LGGSS-----SPSQTLISLGTGNPQGPFPPTQLS----- 1458
QY 1401 ALGPLKLPKPAHEGLVATVKEAGRSIHEIPREELRHTEPELPLAPR-PLKEGSIITGGTPLY 1459
Db 1459 -----LTPA--SSLVPT---PAOTLSLAPGPPGLPTQTLAPAPPLAPASPVGPAPHT 1508
QY 1460 DTGASTTGSKK---HDVRLSICSGRTFFPVVPLDVMADARALERACYEESLKSRCPTA 1515
Db 1509 LTLAPASSASLAPASVQTLTUSPA-----PVPTLGPAAQTLALAPASTQSPASQA 1561
QY 1516 SSSGGSIARGAPVIV-----PELGKPRQSPLTYEDHGAPFAGHLPRGSPVTMR 1563
Db 1562 SSLVVSASGAAPLPTVMVSRPLPVSKDEPDTLRSGLPPSPSTATSEGGPRPRQP---- 1617
QY 1564 EPTPR-----LOE-----GSLSSKASQDRKLTSTPREIAK-- 1594
Db 1618 PPPRSPFYLDLSLEKKRQRSELERIFOLSEAHGALAPVYCTEVLDFCTLQPVASPI 1677
QY 1595 SPHSTVPEH-----HPHPIPSVEHLRLGSGVDLYRSHIPLAFDPTSTPRGIPLD 1644
Db 1678 GPRSGPSHPTFTWYTEAHRAVLFPQORLDQLSIEIERFVMP-----PVEAPPP 1729.
QY 1645 AAAAYLPRHLAP-----NPTYPHLYPPYLI-----RGYPD-----TA 1677
Db 1730 SLHACHPPPLWLAQRAAFQOLASELWPRAPLHRIVCNMRTOFPDLRLQYDCGKLOTL 1789
QY 1678 ALENRQ-----TIINDVITQ-----QMHNNTATAMAQADMRLRG----- 1712
Db 1790 AVLLROLKAEGRVLIPTOMTRMLDVLQELTYHGLYLRDGSSTRVEQRAALMERFNAD 1849
QY 1713 -----LSPRESSLALNAAAGPRGIIDLSQVPLVLPVPTPTGTPATAMDR----- 1757
Db 1850 KRIFCILSTRSGGVNLTGADTVVFDSDW-----NPTMDAQADRCHRTGQTR 1900
QY 1758 -----LAYLPTA-----POPFSSRHS 1773
Db 1901 DVHIYRLISERTVEENILKANKQMLGDMAIEGGNFTTAYFKQOTIRELFDMPLEPSS 1960
QY 1774 SS-PLSP-----GGPHTLTKPTTSSSERDR----- 1800
Db 1961 SSVPSAPEEETVASKQHILEQALCRAEDEDIRAAQKAEQVAELAEFNENDGFPA 2020
QY 1801 -----DRDRDREREKSLTSTTVEH----- 1823
Db 2021 GEGEAGRPGAEDEMSRAQETAAALVEQLTPIERYAMKFLASLEVSREELKQAEQV 2080

QY 1824 -----APIWR-PCTEQSSSGS--SSGGGGSSSRSPASHSHAHQH--SPISPRT 1867
Db 2081 EAARKOLDQAKBEVFLPQEEBEGPGAGDESSCGTGGTHRRSKAKAPERPGTRVSERL 2140
QY 1868 QDALQORPVSRLNTHGKGIITAVEPSKPTVLRSTST---SSPVRPAATFPFATHCPGCGT 1924
Db 2141 RGARAEETOAGNHTP-----VISAHQ-----TRSTTTTPRCSPARERVPRAPRPTPAS 2190
QY 1925 LDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFIAKAPPARSGLEBPASSPSKSEPRPL 1984
Db 2191 APAAIATL---VVPVVSAPVPIAPNPITLPHVILPSPPPPSQIIPCSSPA--CTPPPA 2245
QY 1985 VPPVSGHATARTPAKNLAPHASP---DPPAPPASAS-----DPHREKTQSKPFSI 2033
Db 2246 CTPPPAHTP---PPAQTCLVTPSSPLLLGPPSPVISASVTNLPLGLRPEALCAQALASP 2302
QY 2034 QEELSLGLVHGSS-----YSPEGVEVS-----PV-----SSPSLTHDKG-LPKHLE 2075
Db 2303 ESLELASVASSETSSLSLVPKDLLPVAVEILPVSSEKNLSLTPSASLTLLEAGSIPNGQE 2362
QY 2076 ELDKSHLEGELRPKQPGVKLGCEAAHLP-----HLRLPESQPSSSPLLOTAPGVKGH 2129
Db 2363 QEAPDSAGETTLTLP-----EGEELPLCVSESGLELPPSAASDEPLOEPLADR-- 2413
QY 2130 QRVVTLAHISEVITQDYTRHHHPQOL-SAPLPAPLYSFPQASCPLVLDLRRPPSDLYLPPP 2188
Db 2414 -----TSEELTEAKTPTTSPEKQELVTAEVAAPSTSSSATSSP----- 2452
QY 2189 DHGAPARGSPHSGGKRSPENKTSVLGGEDGIBVSPPEGMTEPGHRSASVYLLYRD 2248
Db 2453 -----EGPSAPRPP-----RRT 2465
QY 2249 GEOTEPSPMGSKSPGNTSOPPA--PFSKLTESNSAMVSKKQOEINKLNTHNRNEPEYNI 2306
Db 2466 SADVEITRGQGTGRPG---QPPGPKVLKLPQRLVTVEEKELVQRRR----- 2509
QY 2307 SQPGBIFNMPAITGTGLMTYRSQAOVEHASTNMGLEAIRKALMGKYDQWEESSPLS-- 2364
Db 2510 QQRGAASLTLPVGSET-----SASPGSPSV--RMSGP---ESSPIGGP 2549
QY 2365 -----ANAFNPLNASALPAAMP---ITAADGRSDHTLT 2395
Db 2550 CEAPSSSLTPPQQPFIAARRHIELGVTGGSPENGDGALLAITPPAVKERRGRPPKKNR 2609
QY 2396 SPGGGKAKVSGRPSRKAAS---PAPG-----LASGDRPPSVSSVHSEG 2437
Db 2610 SPADAGRQVDEAPSTLKGTNGADVPGPETLIVADPVLQPIPGQPLGQPVH--- 2666
QY 2438 DCNRRTPLTNRVME---DRPSSAGSTPPFPYNPLIMRLQAGVMA-------PPPPGLP 2485
Db 2667 ---RPNPLLSPEVKERRRRGRPPKARDLP--GTISSAGDGNSESRTQPPHP 2713
QY 2486 AGSGPLAGPHHAWDEEPKPLLC 2507
Db 2714 SPLTLP-----PPLLVLC 2725

RESULT 51

ABR42219

ID ABR42219 standard; protein; 2193 AA.

XX

AC ABR42219;

XX

DT 28-JUL-2003 (first entry)

XX

DE Human protein kinase 85924.

XX

KW Human; protein kinase; 85924; enzyme; gene therapy; cytostatic.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Modified-site 6. .11 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 31. .34 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 35. .38 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 42. .47 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 67. .69 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 136. .138 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 143. .148 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 154. .157 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 154. .156 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 174. .177 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Domain 181. .439 /note= "protein kinase domain"
FT /note= "protein kinase domain"
FT Modified-site 190. .195 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 191. .193 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 203. .206 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 215. .218 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 219. .221 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 250. .252 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 252. .255 /note= "Amidation"
FT /note= "Amidation"
FT Modified-site 267. .272 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 268. .270 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Active-site 305. .317 /note= "serine/threonine protein kinase active site signature"
FT /note= "signature"
FT Modified-site 325. .325 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 333. .335 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 335. .338 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 393. .396 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 398. .403 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 422. .428 /note= "tyrosine kinase phosphorylation site"
FT /note= "tyrosine kinase phosphorylation site"
FT Modified-site 454. .457 /note= "Amidation"
FT /note= "Amidation"
FT Modified-site 456. .459 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 492. .495 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 517. .520 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 517. .519 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 600. .603 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 604. .607 /note= "glycosaminoglycan attachment site"
FT /note= "glycosaminoglycan attachment site"
FT
FT Modified-site 605. .610 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 625. .628 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 746. .751 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Region 774. .795 /note= "Leucine zipper signature"
FT /note= "Leucine zipper signature"
FT Modified-site 800. .805 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1064. .1069 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1074. .1079 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1079. .1082 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1079. .1081 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1089. .1094 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1106. .1109 /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT /note= "cAMP/cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 1108. .1110 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1113. .1116 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1149. .1151 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1179. .1182 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1188. .1202 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1204. .1209 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1218. .1223 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1221. .1224 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1242. .1245 /note= "Amidation"
FT /note= "Amidation"
FT Modified-site 1242. .1244 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1288. .1291 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1288. .1290 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1332. .1337 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1339. .1342 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1355. .1360 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1362. .1365 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1386. .1391 /note= "N-myristoylation site"
FT /note= "N-myristoylation site"
FT Modified-site 1398. .1401 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1398. .1400 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1463. .1466 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1467. .1470 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1482. .1484 /note= "protein kinase C phosphorylation site"
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 1485. .1488 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1508. .1511 /note= "casein kinase II phosphorylation site"
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 1523. .1525 /note= "RGD cell attachment site"
FT /note= "RGD cell attachment site"
FT

FT	Modified-site	1533..1538	/note="N-myristoylation site"	QY	434	EOKETFRKFMHOPK-----NFGLIASFLERKTVAECVL-----YYLTQKQENYKSLV	483
FT	Modified-site	1547..1549	/note="protein kinase C phosphorylation site"	Db	316	-----FITGPTGSKVIGDGL--ATLKASPAKSVIGTPEFMAPMEYEEHDES	363
FT	Modified-site	1573..1578	/note="N-myristoylation site"	QY	484	-----RRSYRRGKSQQQQQQQQQQQQQQQ-----MPSRQSEKEKEKEKEKAE-----K	531
FT	Modified-site	1577..1580	/note="casein kinase II phosphorylation site"	Db	364	DVYAFGCMLEMATSEYPYSECQNAAYIKRVKTCGPKASFEKVHDPETIKIIEGECICKN	423
FT	Modified-site	1582..1584	/note="protein kinase C phosphorylation site"	QY	532	EEKPEVENDKEDLLKEK--TDDT-----SGEDNDEKAEVASKGRTANSQCRKRGRIT	583
FT	Modified-site	1622..1625	/note="casein kinase II phosphorylation site"	Db	424	KEERYEI-----KDLISHAFADTGVRLAEEDHGKSTIAL--RLWVEDPKLKGPKK	477
FT	Modified-site	1622..1624	/note="casein kinase II phosphorylation site"	QY	584	RSMAEANSE-EAITPQSAELASMEINSSRWTEEMETAKKGLLEHGHENWASAIARMVG	642
FT	Modified-site	1626..1631	/note="protein kinase C phosphorylation site"	Db	478	DNGAIBFTFDELEKETPDEVAQ---EMIESGPFHESDVKIVAKSI---RDRVALIQW--	527
FT	Modified-site	1632..1635	/note="N-myristoylation site"	QY	643	SKTVSOCKNFYNYKQKQNLDELLOHKLKMEKERNARRKKKAPAAASEEAF-----	696
FT	Modified-site	1642..1647	/note="casein kinase II phosphorylation site"	Db	528	-----RRRIWIPALQ---KEQDVGSPDKARPPVPVPLQVQVTHAQAGQ	569
FT	Modified-site	1661..1663	/note="N-myristoylation site"	QY	697	--PVVVEDEMEAS-----GVSNGEEMVEEAELHASGNEVPRGECSPATVNNSSDT	748
FT	Modified-site	1661..1663	/note="protein kinase C phosphorylation site"	Db	570	PGPP--EPPEADQHLPTLPTLSATSLASDSTFDSGGSTVYSQSSQQSVMGLSLA	627
FT	Modified-site	1685..1688	/note="casein kinase II phosphorylation site"	QY	749	ESIPSPHTEAAKDTGQNGPKPATLGADG-----PPPGPPTP--PRRTSRAPTEP--	796
FT	Modified-site	1697..1699	/note="protein kinase C phosphorylation site"	Db	628	DRAPSPAQCVCSPVSEGVPLQSLPSLGAYQQTAAPPPLAQPTPLPQLVLAQPVVPLQ	687
FT	Modified-site	1713..1716	/note="casein kinase II phosphorylation site"	QY	797	-----TPASEATGAP-----TPPPAPPSAP-----	818
FT	Modified-site	1728..1731	/note="casein kinase II phosphorylation site"	Db	688	PVPPHLLPYLAPASQ--VGAPALQKLPQAPLQAPLQAPVPPMPPIPVVPPITPLAGIDG	746
FT	Modified-site	1742..1745	/note="casein kinase II phosphorylation site"	QY	819	--PVVPEKSEKEETAAAPVVEGE-----EQKPPAAEELAVDTGKAE-----P	861
FT	Modified-site	1763..1768	/note="N-myristoylation site"	Db	747	LPLALP-----DLPTATVPPVPPPPQYFSPAVILPSLAAPLPPSPALPQVAKLPHPPGAP	802
FT	Modified-site	1771..1774	/note="CAMP/cGMP-dependent protein kinase phosphorylation site"	QY	862	VKSECTEEAEFGPAKGKDAEAAEATAEGALKAEKGGSGGRATTAKSSGAPQSDSATS	921
FT	Modified-site	1815..1818	/note="casein kinase II phosphorylation site"	Db	803	LAMPCHTIVNPAP-----ATIPLLAVAPGVAALSIH	834
FT	Modified-site	1817..1820	/note="Asn is N-glycosylated"	QY	922	SADEVDEAEGGDKNRLSPRSLTPTGDP-----RANASPKPLDLKQKQRAAI	973
FT	Modified-site	1870..1873	/note="casein kinase II phosphorylation site"	Db	835	SA--VAQLGQVQYPAAPFQ---MAPTDVPPSPHHTVQNMRAFPQP-----AL	878
FT	Modified-site	1917..1920	/note="casein kinase II phosphorylation site"	QY	974	PTQVTKVHPREDAAATKPAAPPAPPQNLQSPESDAPQPGSSPRGKSRSPAPPADKE	1033
FT	Modified-site	1970..1973	/note="casein kinase II phosphorylation site"	Db	879	PP-----OPTLPQVPLPQPTLPFPVLPQP-----TRPQVLPQPQ	917
FT	Modified-site	2017..2020	/note="casein kinase II phosphorylation site"	QY	1034	AFAAEAKLPGDPPCWTSGLPFPVPPPREVIKASPHADPSAFSAPP-----GHPLPLGL	1088
FT	Modified-site	2017..2020	/note="casein kinase II phosphorylation site"	Db	918	PMLPPQVLPQP-----ALPVRPEPL--QPHLPEQAAPAAATPGSQILLGHAPYAV	967
FT	Modified-site	2057..2060	/note="casein kinase II phosphorylation site"	QY	1089	HDATAR--PVLPRPP--TISNPPPLISAKHPSVLEIROIGAIQ--GMVQLHVHPYSEHAKAP	1144
FT	Modified-site	2062..2065	/note="casein kinase II phosphorylation site"	Db	968	DVAAQVTVFVPPAAVLSPLPEVLLPAAPELLPQPFSSLATVSASVQSVPQTATLLPP	1027
FT	Modified-site	2067..2070	/note="casein kinase II phosphorylation site"	QY	1145	VGPVTMGLPLPMDPKLAP-----FSGVKQSLSPRQAGPESLGVPTAEASVLRG	1197
FT	Modified-site	2072..2075	/note="casein kinase II phosphorylation site"	Db	1028	ANP-----PLPGGPGIASPCPTVLTVEPVVEEQAS---QDKPP---GLPQSC	1070
FT	Modified-site	2077..2080	/note="casein kinase II phosphorylation site"	QY	1198	TALGSPVGGISITKGIPSTRVPDSAI--TYRGSITHGTADVLKYGKTTITRIIGDPSRLD	1256
FT	Modified-site	2082..2085	/note="casein kinase II phosphorylation site"	Db	1071	-----SYGSDVTSG---KELSDSCGAFGGGRLEGRAAKHRRS--TR-----ARSRQ	1116
FT	Modified-site	2087..2090	/note="casein kinase II phosphorylation site"	QY	1257	RGRDLSLPGHVIYEGKKHVLISYEGGSMVTCQSKEDGRSSSGPPHETAAPKTYDMEG	1316
FT	Modified-site	2092..2095	/note="casein kinase II phosphorylation site"	Db	1117	RASRPRLTILNCVTGDK-----MVEQLE-----TNHKNMVFKEFD--LDG	1156
FT	Modified-site	2097..2100	/note="casein kinase II phosphorylation site"	QY	1317	RVGRAISSASISGLMCGRAIPPERHSPHLKEQHHRGSIQTGIPRSYVQAQEDYLAREAK	1376
FT	Modified-site	2102..2105	/note="casein kinase II phosphorylation site"	Db	1157	DAPDEIATYWE-----HDFILQERETFIQMDKMDKAE	1193

Query Match	3.4%;	Score 451.5;	DB 6;	Length 2193;	
Best Local Similarity	20.2%;	Pred. No. 3.9e-16;			
Matches	548;	Conservative 304;	Mismatches 979;	Indels 881;	Gaps 141;
QY	55	PGSIIQORRRPSSLSEFQPNERSQELHLRPSHSLYLPGLKSEMEFTESKPRLELL-	113		
Db	9	PAGMAEPRAKAA-----RPGQRF-----LRSVSVESDQDEPPGLEAAE	47		
QY	114	-----PDPLLRPSPLL-----ATQPAGESEDLTORSUTGKLEPVSPSPPHTD	157		
Db	48	APGPPQPLQRRVLLCKTRRLIAERARGPAPAPAAAL-----VAQPGAGAPADAG	101		
QY	158	PE-----LELVP-----PLSKSEELIQNMDRVDRBITMVEQQISKLKKQ	197		
Db	102	PEPVGTQEPGPDIAAAVETAPADGGPR-----EAAATVRKED	141		
QY	198	QQLBEEAAKPPPEPKVPSPPIESKHSRLVQIYIDENRKKAAEAHRLLEGLGPOVELPLY	257		
Db	142	--EGAAEAKPEGRTRRDEPEEED-----DEDDLKAVAT-----	174		
QY	258	NQPSDTRQYHENIKINQAMRKLILYFKRNHARKQWKQFCORYQLMEALEKKVERIE	317		
Db	175	--SLDGRFLKFDILELGRGSKFTYVKGLDTETWEVAV-----CELQDRKLTLE-----	221		
QY	318	NNPRRAKESKVREYKEQPEIRKQRELOERMOSRGVQSGLSMAARSEHEVSEIID	377		
Db	222	---RQRFKE--EAEMLXGLQHPNIVRFYDFWE-----SSAKGKRCI--VLVT	261		
QY	378	GLSEQENLEKQMRQLAVIPMLYDAQOQRIKFINMGLM-----ADPMKYVKDQVNMNWS	433		
Db	262	ELMTSGTLTKYLRKFRKVMKPKVLRSCRQI-----LKGLFLHTRTPPIIHRDLKCDNI--	315		

1377 LLKRE-----GTPPPPPPSRDL---TEAVKTOALQGL-CLKPAHEG-----LVATVK 1419
1194 MLESDTDADRGSDPGTSPHLSTCGLTGCEESRQANAPVQVNLHKGKWFICPVA 1253
1420 EAGRSIHEIPREELRHTPELPAPLPLKBSITQGTPLKYDGTGASTGS---KKHDVRSLI 1477
1254 E-----HPAP-EAPESSPPLSSLP-PEAS-QDSAPYKQDLSSEKEQPSFLASQQLLSQA 1305
1478 G---SPGRTPFPVHPLVDVADARALERACYESLSKRP---GTASSGGSIARGAPV-IVP 1531
1306 GPSNPEGAPPAPLAPSSPVTALPQDGAAPATSTWPEPASQASQAGGP---GTQGLTS 1362
1532 ELGKPRQSPLTYEDHGAFFAGH-LPRG-SPVTWREPTPLQEGSLSSSSASQDRKLTSTP 1589
1363 EL---ETSQPLA-ETHEAPLAVQLVGLAPCT---PAPE-----AASTRDASAPREPLPP 1412
1590 REIAKSPHSTVPE---HHPHPTISPYEHLRLGVSVDLYRSHIPLAPDPTSIIRGIPDLDA 1646
1413 AP-EPSPHSGTQPALGQAPLPLP-----AAGVAVSLATSQLP-----SPPLG----- 1454
1647 AAYLPRHLAPNPTYPHLYPPYLIR-----GYPDTA-ALEN---RQTIINDYITS 1692
1455 -----PTVPP-QPPSALESDEGPPRPVGFVSTIKSLDEKURTLTYQSHVT 1501
1693 QQMHNHTATAMAQR---ADMRLGLSPRESSLALNYAAGPRGIIDLQVPHLPVLVPPTPG 1749
1502 SSASACTPVEVDGDRDFTLEPLRGDQR-----SEVCGDLALPPVPK 1543
1750 TPATMDRLAYLPTAPQPPSSRHSSPLSPGGTHLTKTPTTSSSRER---DRDRERDR 1806
1544 EAVSGRVQL-----PQLVKSSELAFTRGAVMEQGTSSMTSESSPRMLGYDRD----- 1592
1807 DREKXSILTSITTVHEAP-----IWRPG-----TEQSSGSGSS-----GGGSGSS 1848
1593 -----GRQVADSHVPSVQDVPAPFVRPARVEPTDRGGEAGESSAEPSPDMGTVGQA 1648
1849 SRPASHSHAHQHSFISPTQDALQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSPVR 1908
1649 SHPQT-----LGARALGSPKRP-----EQQDVSSPAK 1676
1909 PAATFPATHCPGGTLDGYYPTLMEPVLLPKEA-PRVARPERPRADTGHAFKAPKAPAS 1967
1677 TVGRFSVWS-----TQDEW---TLASPHSLRYAPPDVVLDAPSPDVKLAVRRAQTAS 1728
1968 GL-----EPASSPKSGSERPLVP-----PVSGHAT---TARTPAKNAHPHA---SP 2009
1729 SIEVGGEFVSSSDGEGPARPVPVQKASLPVSGSVAGDFVKKATAFQPSRAGSLGP 1788
2010 DPPA-----PPASADPHREKTSQKPFISIQELESLGLYHSGSSYSPGVEPVSPVSPS 2063
1789 ETPSRVGMKVPTISVTSFHSQSSYISDNDNDELE-----1822
2064 LTHDKGLPKHLELDKSHLE-CEIAPKQPGPVKLGGEAAHLPHLRPLP-----2110
1823 ---DADIKKELOSLRKHLKEISELOSQKQEI-----EALYRLGKPLPNNVGVFFHTAPP 1875
2111 -----ESOPSSSPLQATPGVKQHORVTLAQHISEVITQDTRHHPPQLLSAPLAP 2162
1876 TGRRTKTSKLAGKLLN---PLVRQLKVVASSTGHAD-SSRGPAPKAPQAQASVGLTA- 1931
2163 LYSFPGASCVDLDRP-----PSDLYLPPDPHGAPARGS-----PHSE---GGKRRP 2207
1932 ---DSTGLSGKAVTQPCSVRASLSSDICSGLASDGGGARGQGWTVVHTPTSERVTYKSSS 1989
2208 EPNKTSVLGGEGDIEPVSPGWTGPHRSRAVYPLVR---DGEQTEPSRMGSKSPGN 2264
1990 KPARFLSG-----PVS-----VSIWALKRKLGLGKHSKSSSTSSLAGP 2030
2265 TSQP-PAFFSKLTESNAMSVMKSKQKQKINKLNTHRNEPEYNTSQPGTEIFNMPAITGTG 2323
2031 EPGQPQALHVAQVNNNS---NNKKGFTDDLH-----2059
2324 LMYRSQAVQEHASTNMGLEAIITRKALMGKYDQWERSPPUSANAFNPLNASLPAAMPI 2383

Db 2060 -----KLVDWTSKTVGAQL---KPTLNOLKQTKLOQMEQAAG-----2096
Qy 2384 TAADGRSDHTLTSPPGGGKAKVSGRPSRKAKSP-----APCLASGDRPPSVSVHSEGD 2438
Db 2097 WAAAPGEA-RAMTAPRAG-----VGMPLRPPAPGPLSTTVIPGAA-----PTLSVPTPDGA 2145
Qy 2439 CNRRTPLTNRVW 2450
Db 2146 LG--TARRNQW 2155
RESULT 52
ABR41356
ID ABR41356 standard; protein; 2735 AA.
XX ABR41356;
AC ABR41356;
XX 02-JUN-2003 (first entry)
XX Human DITHP transcription factor.
DE Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
XX cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW transcription factor.
OS Homo sapiens.
XX WO200297031-A2.
XX 05-DEC-2002.
XX 27-MAR-2002; 2002WO-USO10056.
XX 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshy SR;
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Feralta CH, David MR, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
XX N-PSDB; ACC46296.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 891; 591pp; English.
PS The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP

Db 1535 TQFPDLRLIYDQGLQTLAVLLRQLKASGRHVLFTQMTMRMLDVLEQFLTYHGHLRL 1594
QY 1698 NTATAVAQADMLRG-----LSPRESLALNYAAGPRGIIDLSQVPHLPVLVPP 1746
Db 1595 DGSTRVEORQALMERFNADKRIFCFILSTRSGGVNLTGADTVVFDSDW----- 1645
QY 1747 TQCTPATAMDR-----LAY 1760
Db 1646 NPTMDAQADRCHRIQGTDRDHYRLISERTVEENILKANQKRMGLDMAIEGNGFTTAY 1705
QY 1761 LPTA-----POPFSSRHSS-PLSP-----GGPHTLTKPTTSSSERDR----- 1800
Db 1706 FKQOTIRELFDMLEPSSSSVPSAPEEEETVASKQTHILEQALCRADEEDIRAATQA 1765
QY 1801-----DRDRDREREKXSLTSTTVEH----- 1823
Db 1766 KAEQVAELAEFNEGDFPAGEGEAGRPGAEDSEMSRAEQEIAALVEQLTPIERYAMKFL 1825
QY 1824-----APIWR-PGTEQSSGSG--SSGGGGSSSR 1850
Db 1826 EASLEVSREELKQAEQVEAARKDLDQAKEVFRLPQEEERGPGAGDESSCGTGGTHR 1885
QY 1851 PASHSHAHQH--SPISPRQTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST--SS 1905
Db 1886 RSKKAKAPERPGTRVSERLRGARAEQTGANHTP---VISAHQ-----TRSTTTPRCSS 1935
QY 1906 PVPAATFFPATHCPGLGTLGDVGYPTLMEFVLLPKAPRVARPERPRADTGHAFKAPA 1965
Db 1936 PARERVPRPAPRPRPTPASAPAAIPAL---VPVFSAPVPSAPNPITILPVHILSPSP 1992
QY 1966 RSGLEPASSPSKSGSERPLVPVSGHATTIARTAKNLAPHASP---DPPAPPASAS--- 2019
Db 1993 PSQIPCCSPA--CTPPACTPPATP---PQAQCLVTPSPSPLLLGPPSPISASVTN 2047
QY 2020-----DPHREKTSQFSTIQELRLSLGVHGS-----YSGPEGVPS-----PV----- 2059
Db 2048 LPLGLRPEALCAQALASPESLASVASSETSSLVPPPKDLPVAVELPVSEKNLSL 2107
QY 2060--SSPSLTHDKG-LPKHLELDKSHLEGLPKQPGVKLGGEAHL-----HLRPLP 2110
Db 2108 TPSAPSLTLEAGSIPNGQEQAPDSAGTTLTVLP-----EGEELPLCVSESNGLELP 2160
QY 2111 ESQSSSPLLOTAPGVKGHQRVTLAQHISEVITQDYTRHHPQOL-SAPLPAPLYFPFGA 2169
Db 2161 PSAASDEPQLEADR-----TSEELTEAKTPTSSPEKPOELVTAEVAAPSTSSAT 2213
QY 2170 SCPVLDLRPPSLYLPDPHGPARGSPHSEGGKRSPEPKTSV-----LGGEDGIE 2223
Db 2214 S-----SPGPGSPAR-----PPRRTSADVEIRGQGTGRPG-Q 2245
QY 2224 PVSPPGEMTEPGHRSNAVYLLYRDGEGQTEPBRMGSKSPGNTSQ-PPAFSKLTESNSAM 2282
Db 2246 PGPGKVLRLPGLRVTV-----EKELVRRRQQRGAAGTLPVGVSETSASPGSPS 2297
QY 2283 VKGKK-QEINKNLNTHRNPEYNIOPGTEIFNMP-----AITGQ-----LMTYR 2328
Db 2298 VRMSGPSSPPGPGCEAAPSSSLPTPQOPFIARRHIELGVTGGSPENGDAALLAIT 2357
QY 2329 SQAVQEH-----ASTNGLEAIIRKALMGKYDQWESP-PLSANAFNPLNASAS 2376
Db 2358 PPAVKRRRRGPPKKRNSPADAGRGVDEAPSLTKGTNGADVPVGPETLIVADPVLPEQL 2417
QY 2377 LPAAMPITADGRSDHTLTSPGGGKAKVSGRPSRRKAKS-PAPGLASGRDRPSPSVSVHS 2435
Db 2418 IPGQPLGPGQVHRNPPLLSP---VEKRRGRPP--KARDLPIG-----TISGA-G 2463
QY 2436 EGCNRRPTLTNRWEDRPSAGSTPPFPYNPLINRLQAGV-----MASPP----- 2480
Db 2464 DGNSESRT-----QPPHPSPILTLP--PLLCVPTATVANTVTTVTISTSPPKRGR 2514
QY 2481 PGLPAGSGP 2490

Db 2515 PPKNPPSPRP 2524
RESULT 53
ABB68397
ID ABB68397 standard; protein; 2768 AA.
XX
AC ABB68397;
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31983.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12500.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX
PS Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2768 AA;
Query Match 3.4%; Score 450.5; DB 4; Length 2768;
Best Local Similarity 20.6%; Pred. No. 6.1e-16;
Matches 498; Conservative 264; Mismatches 944; Indels 715; Gaps 117;
QY 504 QOQOQOQMPR---SSQEEKDEK-----EKEKEA-----EKEKE-KP-EYENDK 542
Db 478 EEEEEKPTPAEGSGSEEEKVKVTAAPETEDEAKPTSPVASDEKEQEPKPESGGDE 537
QY 543 EDLLKEKTDGTDGDNDEKEAVASKGRKTANSQGRKRGKTRITRSMANEANSEEAITQOQA 602
Db 538 ELDLKPTAPTAGATSASE-----SSEODEGKSTAPTSDVDIEPA-KPTESS 585
QY 603 ELASME-----LNSSSRWTEEMETAKKLLHGRNWSAIARMVSGKTVSQCKNF 652
Db 586 EASGEGEDVAKETTPAGEASIAEGEEI---VKGTTPAGEPSS----- 625
QY 653 YFNYKQRQNLDEILOQHKLKMEKERNARRKKKAPAAASEEAFPPVVEDENEAAGVSG 712
Db 626 -----EGDEEIVKGTTPAESSESSSEDELTKVTTPAGE-----PSVAGEEETAKETTPA 674

QY 713 NEEMVEEABALHASGNEVPRGCS-----GPATVNNSSDTESTIPSPHHTBAKD 761
Db 675 GEPSIAGEEIEIVKT---TPAGESIAGEBEEIVKVTTPAGESSEGEIEIKVTTPA--- 728
QY 762 TGONGKPPATLADGPGPPGPPPP---RRTSRAPITPTPASEATGAPTPPPAPPSP 815
Db 729 -GSSSEGDEEIVKESTPAGEPISEGEEDVIKATTS-----PKSDIEGVKEPETATE-- 780
QY 816 SAPPVVVPEKEEETAAAPVVEGEEQKPPAAEELAVD--TGKABEPPVKSECTEEBAEG 873
Db 781 -----VPABEEVDFAKPTPIABAEIEE-PIAGTPIPTDGISGE-BEIVKGTTPQTLEEQ 832
QY 874 PAKGDAEAAEATBAGALKAKEGSGRATTAKSSCAPODSSATCSADEVDEAGGD 933
Db 833 P-----EISESTEVPVAAEDDLSSTASAIASSTEGVQDAASETTSSA-----PARAGD 882
QY 934 KNELLSPRPSLLPTGDPRANASPOKPLDLKOLKORAAAIPIQV-----TKVHEPPR 986
Db 883 KD-----EAATVPTAODK-----DDEVEQDADLPLVEDVQVSTAKTTTTTEQPK 928
QY 987 EDAAPTCKAPPAPPQNLQOPESDAPQOPGSSPRGKSRSP-APPADKEAPFAABAQKLPGD 1045
Db 929 EES-----STEADAEIEVTTSSPADKQEVPEASPADKOHKDEEDVQTATD 974
QY 1046 PPCWTSGLPPVPREVIVKASHPAPPSAFS-----YAPP-GHPLPLGL-----HDT 1091
Db 975 LPI-KSDIGPVVDTEATTQOPETSDETATDKPPSVLPPVQEVPSSTAKVDRNDFET 1033
QY 1092 ARVLPRPPTISNPPPLISSAKHPSVLEROIAGISQMSVOLHPVSEHAKAPVGVWTMG 1151
Db 1034 EKTLPPSGEDQSSEPL-----PAMDLPAGIPEGDCU-----VEGKTYANNTIVPAT-- 1081
QY 1152 LPLPMPKPLAPPSGVKQEQLSPRGOAGPPSPGVPTAQEASVLRG---TALASVPEGSI 1208
Db 1082 --APCDVSCKICISLVAQQM---ECKLPENLEKCTV-AADLLDCCPTVICDESTESA 1134
QY 1209 TKGIPSTRVPSDAITVRGSI-THGTPADVLYKGTITR-----II 1247
Db 1135 EKDEESTAKPDNKIDEDVSEISTEIEPKOVIMPTGITQEQLSHVKPDEEIQPVTSVPAQF 1194
QY 1248 GEDSPERLDR-----GREDSLPGHVIYEGKGHVLSYEGMSVTOCSKEDGRSSGPPH 1302
Db 1195 DESTTAKVKKPIDESAEDKKPIGE-----SEEDSKPIDE-SEEDKK-----PVE 1238
QY 1303 ETAAPKRTYDMMEGR---VGRAISSASIE-----GLMGRAIPPERHSPHHLK 1346
Db 1239 ESAEDKPVEDSEKEKPLTVIPASEIEKESKEPEDEKTEADFAAPTEQPEATTAQIA 1298
QY 1347 E--OHHIRGSI-TQGIPIRSVVEAQEDVLRREAKLLKREGTPPPP---PPSRDLTBAYKT 1399
Db 1299 DTAEKEVDDKLATTSAPVS---GEDEL-KPADEKKRTETAQIPDAEIPASTDEPESSTE 1353
QY 1400 QALGPLKLPKPAHGLVATVKEAGRSIHEIP-----REELRHTPEL-PLAPRPLKEGSITQ 1453
Db 1354 LPTVDLKKPEEDSTKGTEAPESDKVPEVPTSASTENEIESDKFTTVAPPKISASDETE 1413
QY 1454 GT-----PLKYDTGASTTGKCHDVRSLIGSPGRTFPFPHPLDMADARALERA 1502
Db 1414 PTAEBEDLVPAFTFPISEFEVST---KKPAVQ---GPPLTLAPAQEKKPVDAETSTE 1467
QY 1503 CYE-----ESLKSRPGTASSGGSIGARGA-----PVIIVELGKPROSPLTYEDHG----A 1548
Db 1468 DISTEPSAEVEKASGETSESDEINIDAGASSTVPVPSADEDKTPTSTEKTEADDKFTTVA 1527
QY 1549 PFAG-----HLPR-----GSPVTWREPTPLQEG----- 1572
Db 1528 PLAGDEENLPLKPODIFEEEPAPVAVTTAAPSODGQKQVVEBKPIEDGOKPIEDET 1587
QY 1573 -----SLSGSKAQDRKLTSTPREIAKSPHSTVPEHHPHPISEYELLRGVSGDLVYRSH 1627
Db 1588 STPTSENIEPEPSDRATTIAPKEPSEFSTCAPTKDEPAEFTSDAPESDESKETPESE 1647
QY 1628 IPLAFDPTSIPIRGIPIDAAAAYVILPHLAPNPTYPHLYPPYLRGYPDTAALENRQTIIN 1687

Db 1648 VP-----TTVAPAGEKIPTSS---ITPDEEFTATSAFVAKP-----DEDVEKSTSTIPT 1694
QY 1688 DYTTSQOMHNTAT-----AMQORADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLP- 1741
Db 1695 DAPASSEEDENSTDOIPEVEPEKKETPAQTPEEGDI-VGATAAPTDSDEVPPVQLPE 1753
QY 1742 -VLVP-PTPG-----TPATAMRLAYLPTAQPFSSSHSSPLSPGPGTHLTKP 1788
Db 1754 EVLAEIPQSTETGKIQQDETTAAPSIDRKE--PYVTEIDEEATTVAPISEKD-----EKP 1807
QY 1789 TTTSSSERDRDRDRDRERERKSILTSTTTTVEHAPIWRPGTEQS--SGSSGSGGGGGS 1847
Db 1808 TEEKPVQKPTGEEPESEEEKPIEQDVST--EGPV---STEASEAGSTESS----- 1856
QY 1848 SSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKGIITAVEPSPKTLVRSTSTSSPV 1907
Db 1857 -----EEVKPSTGEVAEKP-----EDKQP-----SSTAQ 1881
QY 1908 RPAATPPP-ATHCPLGGTLDGVYPTLMEPV-----LLPK-----EAPRV--A 1946
Db 1882 APVETIPEISTELP---AQDGDKPTSEAPVDSDEDTSAPEDEKIPSVSGVEGEVETTA 1938
QY 1947 RPERPRADTGHAFAPPARSGLEPASS-----PSKGEPRPLVPVPSGHAT 1993
Db 1939 SPOAEEDE-----LKTPAES--EPSSTDKVPETEXKPEDETKADETPESVTQVSDVAT 1991
QY 1994 IARTPA-----KNLAPHASPDPP-----APPASADPHREKTQSKPFSIQELE-- 2037
Db 1992 STSAPVAGDIEKDEQATTASPEEEIEKPTIAPAAIPOP-----SEKEPVDSQVESGT 2047
QY 2038 ---LRLSLGVHSGSYGEGVEPVSPVSPSLTHDKGL-----PKHLEEL----- 2077
Db 2048 KATPAESDGPIDEIAPATSGPIDEASTAAPTKEESTTVAASAASPAVHDDDKI KDVTTTQP 2107
QY 2078 --DKSHL-----EGELAPKQPGPVKLGGEAAHLPHLAPLPESQP 2114
Db 2108 VADEKEVAAPQDDETKTSIDVSTDSPSTAQDDEKODKTEAPV-----APTTVSSP 2155
QY 2115 -----SSPLLOTAPGVKGHORVV-----TLAQHISEVITQDVTTRHHPOQLSAPL-- 2159
Db 2156 TADSAADSTPTVEVSPVEIDTKPMDDIMSQTIAHTADGAASTEDDEQDQ---APVTV 2212
QY 2160 -----PAPLYS-----PFGASCPVLDRRPPSDLYLPPPDHGAPA 2194
Db 2213 SPQDAEKTVPVPAQSDKTPSSEAPQDADEIPATATPLDDNKIPAT--VAPQTDGCVPA 2270
QY 2195 RGSHPHSEGGKSPENKTSVLGGEGDI-----EPVSPPEGMTFEGHRSASV 2241
Db 2271 TAAPLDE-----DKIQTAAAPLDEBKIPSTAAPLDEKIPAPVSPVVDVPSSEKPAV 2324
QY 2242 YPLLVRDGEQTEPS-----RMGSKSPGNTSQPPAFFSKLTSNSAMVK 2284
Db 2325 SE--YDGESESTPPVHDVETSDPTSDAKLKPTSPATPSPESPATEASIVETAPEL 2382
QY 2285 SK-----KQEKINKKLANHNRNPEYNIQSP--GTEIFNMPAITGTGLMYRSQAVQE 2334
Db 2383 EKEVPEKATEQPELEKETPEKATEQPELEKETPEKATE--QPEL-----EKETPE 2430
QY 2335 HASTNMLGLAIIRKALMGKVDQWEEPPPLSANAFNPLNASASLPAAMPITTAAGRSHTL 2394
Db 2431 KATEQPELEKEVT-----DKATEQPEVDEKTTTP-----EPVVKPSLDSTEDEESV 2477
QY 2395 TSPGGGG---KAKVSGRPSRRKAKSP-APGLASGDRPPSVSVSVHSEGD-----CNRR 2442
Db 2478 ESEESADKKDKNKETEEDTKKHEEPFPAVVSEIPQSEEAIVPTTGHPLFLPHLASSTT 2537
QY 2443 TP--LTVRWED-----RPSAGSTPPFPNPLINMLQAGVNMAS-----PPPPGLPA 2486
Db 2538 TTPAVDDRVGEEDDEENTVKLSSSTTTSTTESPVSAPTSTTTTASQQOQIITPPYGV--- 2594
QY 2487 GSGLAPGPHHA-----WDEE 2501

Db 2595 -----HAPEYDEYDEE 2606

RESULT 54
ABG17148
ID ABG17148 standard; protein; 2759 AA.
XX
AC ABG17148;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17139.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73;
DR N-PSDB; AAS81335.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
FS Claim 20; SEQ ID NO 47507; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (II) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2759 AA;

Query Match 3.3%; Score 441; DB 4; Length 2759;
Best Local Similarity 18.4%; Pred. No. 2.1e-15;
Matches 435; Conservative 300; Mismatches 920; Indels 704; Gaps 104;
403 DOQRIKEFNKGLMADPMKVKYKQVNMNMSEKQETFEKFMQHPKXNFGILASFLERKT 462
95 DQSAIKNI-----ESAKGRDWDNSFLQADGTVAETGYHYVVFALAGS---RKD 141

463 VAECVLYYYLTKNENYKSLVRRSYRRRGKSGQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 522
142 ADDTSIYMFYQKVDN-----SIDSWKNA 165
523 KEKEKEAEKEBEKEVENDEKEDLLKEKTDGDDDEKEDNEKAVASKRKT---ANSQGRKK 579
166 GRVFKDSDFDA-----NDPILKDOTQEWSG-----SATFTSGKIRLRYFTYDSGKH 213
580 GR--ITRSMANEANSEAITPQSAELASM-----ELNESSRWTEEMETAKK 625
214 GKQSLTTAQVNVKSDDTLKINGVEDHKTIFDGGKTYQNVQOFIDEGNYTSGDNHLLRD 273
626 G--LLEHGRNWSAIARMVG-SKTVSQCKNFYNYKKRONLDEILQOHLKMEKER--NAR 680
274 PHYVEDKGHKYLVFEAQHGEQKTDQSGESLFNKAYGGTGNFFQSSKLVKQKVEPMNSV 333
681 RKKKKAPAAAEAEAFPPVVEDEMEASGVSGNEEMVEEAEALHASGNEVPRGCSGPA 740
334 RVTFNIPREAAERLILAQSNQQQLDLGILSVQIE-GEQAINLALAQNRSDQVRNMGPM 392
741 TVNNSDTEIPSPHTEAAKDTGQNGPK-----PP-----ATLGADGPPP--GPPTPPRRT 789
393 GAGNSVRMEA-GFPMASGFIIRMNPNPATVMIPPGNGVSSSMMAFGPNPELQPTPRPAS 451
790 SRAPTEP-----TPASEATGAPTPPAPSP-----SAPPVVPVKEKEEETAAAPV 837
452 QSDAMDPLLSGLHIQQQSHPSGLAPPHHPMQFVSVNRQMNPNANPPOLOQQQQQQQQQQ 511
838 EGEBEQKPPAAEELAVDTGKAEPEVKSECTEAEBSGPAKDAEAAEATAEALKAKEKE 897
512 QQQ 571
898 GGSGRATTA---KSSGAPQDSDSSATCSADEVDEAEAGGDKNRLSPRSLTPTGDPAN 954
572 GSLGTWTANQWKKAPLP-----GPMQQLQARPSLAT-----604
955 ASQPKPLDLKQLKQRAAIPTQVTKVHPREDAAPTKAPAPAPPPQNLQPSDAPQ 1014
605 -----VQT-----PSHPPPPPPPGSQQAQAHTN 628
1015 -PGSPRGKSRPAPPADKEAFAABAQKLPDPPCWTGSLPPPPVPPPREVIKASHAPDPS 1073
629 FQMNPNQOFTAP-----QWKSLOGP-----SRVPTPQQPHLTNKP-ASSFS 672
1074 AFSY-APPGHPLPLGLHDTARVLPRLPPTISNP-----PPLISSAKHPSVLERQIGALS 1126
673 SFQGSPPASSPT---VNTQQQMGPRPPO-NNPLPQGFQGVSSPGRNPMV---QQGNVP 725
1127 QGMSVOLHVPYSEHAKAPVGVMTMGLPLPMDPKLAP-FSGVKQEQLSPRGQAGRPESLG 1185
726 PNFVMWQQQPPNQ-----GQSLHPLGLGMPKRLPPGFS-----GQNFPMQ 770
1186 -VPTAQEASVLRGTALGVSFGGSIITKGIPTSRVPSDAITYRGSITHTGTPADLYKGTIT 1244
771 QVP-----STATTTPGNS---GAP-QLQANQNVQHAGGQAGPQNM-----809
1245 RIIGEDSPRLDRGREDSLPKGHVIVYEGKGHVLSVEGMSVTCQCKEDGRSSGPPHET 1304
810 -----QVSHGPPNMWQPSLMGIGHNNNQAGTSGVPQVNLNMQGQPOQGP---857
1305 AAPKRYDMMGEVGRVAISSASIEGLMGRAIPPERHSPHLLKEQHHRIGSITQIGIPRSYV 1364
858 -----SQLMGHQQLV-----SQQMVGQ-----QGTINPQNPMILS 890
1365 EAQEDVLRREAKLL---KREGTPPPP-----PPSRDLTE-----AYTKQALGP-----LK 1406
891 RAQ---LMPQGMVNVPPSQNLGSPQRTTPPKQMLSSQQQPMWAPHNQMMGQGVLLQ 947
1407 LKPAHGLVATVKEAGRSIHEIPRE--LRHTPELPIAPPLKEGSIQTGTPLKYTGAS 1464
948 QNPMIEQIMTNQMGKNQKQNTQNSQNVMPGPAQIMRGFTPNMQGNVQPTGOMSGMPL 1007
1465 TTGSKKHVDYSLIGSPRTFPVHPDLMADARALERACVEESLKRPGTASSSGGSIA 1524

Db 1008 QCGPVNNSPQVNGIQGVLRPPGSPHMAQQH-----GDPATTANNVDSLSQ 1055
QY 1525 GAPVI-----VPELGKPRQ-----SPLTYEDHG-----APFAG-----HLPR 1556
Db 1056 MFEDVSIQTNVPPHVOAQNSASGHSFGMSFNAPFSGAPNGNQMSCQNGPFPV 1115
QY 1557 GSPVTMRBP--TPRLQEGSLSSKASQDRKLTST-----PREIAKSPHS--TVPEHH 1604
Db 1116 NKDVTLTSLPLVNLQSDISAGHFGVNNKQNTNANKPKKKPKPKCKNSQQDLNTPDTR 1175
QY 1605 PHISPVEH--LLRGVSGVDLYRSHIPLADPTSIPIRGI-----PLDAAAAYLP-----RHLA 1656
Db 1176 PAGLEARDQPLPCEQGINLDNSG--PKLPEFSNRPPGYPQSPQVPEORPLQOMPPQLMQHVA 1234
QY 1657 PNTYPHLYP--PYLIR-----GYPTAALENRQTIINDYITSQQMHNTATATAQRAQ 1708
Db 1235 PPQPQQQPOQLPQQQQPPPPSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 1294
QY 1709 MLRG--LSPRESSIALNYAAGRGIIIDLSQVPHLPVLVPPTP-----GTPATAMDR 1758
Db 1295 PPRGLNPDORMPMQSGSVPMVSLQ-----PASVPPSPDKQRMMPVNTPLGNSRK 1350
QY 1759 AYLPTAQ--PFSS-----RHSSPLSGGPHLTKPTTSSSERERDRDRDR 1806
Db 1351 MVYQSPQNPSSPLAEMASLPEASGSEAPSVPGPNMPSHV-----1393
QY 1807 DRERKSILNTTVEHAPIRWRCQESGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSS 1865
Db 1394 -----VLPQNQLMTGP--KGGPSPLATQATPOOPVNSLPSHGH-----HFNVA 1441
QY 1866 RTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST--SPVRPAATFPFATHCLPGT 1924
Db 1442 PTQTS--REKTNRASPRVYPTNNRPSTEPSLSLSPER-----LNAS 1486
QY 1925 LDGVYPTLMEPVLLPKAPVARPERPRADTGHAFIAKPPARSGLBPASSPSKGSBPRL 1984
Db 1487 IAGLFP-----POINILPPLPNLRGF-----DOQGLNPTTLKATGOAPS 1529
QY 1985 VPPVSGHATARTPAKNLAHPHASDPDPAPPASADPHREKTKOSKPFISQLELSRSLGYH 2044
Db 1530 TMNPSNPAI-----PQTHKLSVVVN-----S 1551
QY 2045 GSSYSPEGVPSPVSPSPSLTHDKGLPKHLELDKSHLEGELRPKOPGPVKLGEEAHL 2104
Db 1552 GKQNSGATKRAQSPNS-----RRSPGSSR-----1577
QY 2105 HLRLPESQSPSPLLTQAPGVKHQVVTQAHISEVITQDYTRHHHPQOLSAPLPAPLY 2164
Db 1578 KTTSPGRQNSKAPK-----TLASQTNAAALQ-----1605
QY 2165 SFGASCPLDLRRPPSDLYLPPDHGAPARG--PHSEGGKRSPENKTSVLGGG---ED 2220
Db 1606 -----NVELPRNLVSPPLANPPVPFGFPNNSG--LNPQNTSVSVAAGVGVED 1653
QY 2221 GIBPVGPP-----EGMTEPGHSRAVYP--LLYDRGEQTESPSRMGSKS 2261
Db 1654 NKESLNVPOSDCQNSQSRKEQVNIELKAPAEVKXVVEDQSKDQGPSPDNKPLPS-- 1711
QY 2262 PGNTSQPPAFSKLTESNAMSXKQEIKNKLN--THNRNEPEYNISQPG--TEI--FNMP 2317
Db 1712 -----VEENKNLVSPAMREAPTSLSQLLDNSGAPNVTIKPPGLTDLEVTTP 1757
QY 2318 AITGTGLM-----TYRSQVQEHASTNMGL-----EAILRKALMGKYDWEES---- 2360
Db 1758 VVSGEDLKASVITPTLDLSSSKPEPSNLSLPHSNLCSLVLHPHPELSESSNVAPSTPP 1817
QY 2361 -----PPLSANAF--NPLNASASLPAAMPITTAADGRSDHTLTSPGGGKA 2403
Db 1818 MSRPVSSSISTPLPQNQITVFTSNPITTSANTSAALPHLOSALMSTVVTWPNAGSKV 1877
QY 2404 KVS-----GRP-----SSRKAKSPAPCLA--SGDRPPSVSSV-- 2433

Db 1878 MVSEGOSAAQSNARPOFITPVPFINSSIIQVMKGSQSPSTIPAAPLTTNSGLMPSPSVAVG 1937
QY 2434 --HSEGDGN--RRTPLNRVMDRPPSAGSTPPFY-----NPLIMRLQAGVNASPPPPGLPA 2486
Db 1938 PLHIPQNIKFSSAPVP-----PNAISSSPAENIQTRPLVLSRATPVQLPSP---PC 1987
QY 2487 GSGPLAGPHHAWDEEPKPL 2505
Db 1988 TSSPVV--PSHPVPVQVKEL 2005
RESULT 55
ABB58514
ID ABB58514 standard; protein; 2897 AA.
XX
AC ABB58514;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2334.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers BW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL02617.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 2334; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutic and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2897 AA;
Query Match 3.3%; Score 440; DB 4; Length 2897;
Best Local Similarity 18.8%; Pred. No. 2.5e-15;
Matches 526; Conservative 316; Mismatches 1054; Indels 904; Gaps 119;
QY 60 QPQRRRPSLLS-----EFQPNERSQELHLRPHSHSYLPELGKSEMEF 102
Db 326 QQKKHPTLLDTKRIKKPSRYESTKDEIBEESGEDEEE---BDEEDSELDE--GEDE-ED 380
QY 103 IESKRPRLELLPD-----PLLRPSPLLATQPGASEDLTKDRSLTGKL-----EP 147
Db 381 VDSAAADIEDDEEDTGTGPIEDPT--IVVQGGSGLDCEAIPYNLIGNFDYEDDLEIGA 439

Db 2341 LVPSSADLAR-----FYQVANQOPI-PAVFCRS-----PSSTGSPRH----- 2379
QY 1923 GTLDGVYPTLMEVLLPKAPRVARPERPRADTGHAFKAPPARSGLEPASSPSKSGSEPR 1982
Db 2380 -----LLRQMPGGLPP-----PHASLRPTYGPPPLRGSGPTSTPSTTNSR 2423
QY 1983 PLVPPVSGHATARTPAKNAPHASDPDP-----APPASADPHREKTQSKPFSI 2033
Db 2424 PAY--LHG-----AEHGGSPGPMGGVSSGPPPARHATPHLNPVRAPPI-- 2467
QY 2034 QELESLGHHGSSYSP-GEVPEVSPVSPSLTHDKGLPKHLELDKSHLEGELRKPQG 2092
Db 2468 -----YGNPNYSPLRGAGPT-----GSMR-----PG 2489
QY 2093 PVK-LGGEAAHLP-----HLRPLPSQPS-----SPLLQTAGVKHQVVTLAQHISEVIT 2144
Db 2490 ADVYAGPRGYSYGYPPPPPLSTSAHATSVIVSAP-----HTLTPTNHSVPTLT 2543
QY 2145 QDVT--RHPHQQLSAPLAPLYSPGASCFLV-----DLRRPPSLDLYLPPD-----HGAPA 2194
Db 2544 HGKTPPQQTQSSGPPA-----AAPPPTITSETSSHKPPLASVITSKLTLEAYPI 2597
QY 2195 RGSFHS-----EGKRSPEPNKTSVLGGEDGIEPVSPGEMTEPGHRSASVYPLLY 2246
Db 2598 RKSPIAVADVSGPAEPTRSPADIAEDSGSAHDTRAPSSA-----TGTAUVGFS 2648
QY 2247 RDGEQTEPSRMGSKSPGNTSOPPAFTSKLTESAMVSKKOBINKKLANTHNRNEPEYNI 2306
Db 2649 GSGNAWAHCTGPTPRELQ-----SKL-ELRQOSKLEREQEPSKL-----EEL 2694
QY 2307 SQFGTEFINMPTGTGLMYRSQAVQERHASTMNGLEAII 2346
Db 2695 EQOSKLERAQEREGRGREVTNRELAQEQSKLELGOEQSI 2734

RESULT 56

AAE13839
ID AAE13839 standard; protein; 4019 AA.
XX AC AAE13839;
XX DT 26-FEB-2002 (first entry)
XX DE Human lung tumour-specific protein SCC2-29.
XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
XX KW antisease-therapy; vaccine; immune response; lung cancer; SCC2-29.
XX OS Homo sapiens.
XX PN WO200172295-A2.
XX PD 04-OCT-2001.
XX PF 28-MAR-2001; 2001WO-US009991.
XX PR 29-MAR-2000; 2000US-00538037.
XX PR 05-JUN-2000; 2000US-00589937.
XX PR 18-AUG-2000; 2000US-00640878.
XX PR 22-SEP-2000; 2000US-0234517P.
XX PR 01-NOV-2000; 2000US-00704512.
XX PR 14-DEC-2000; 2000US-00738973.
XX PR (CORI-) CORIXA CORP.
XX PA Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
XX PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX PI WPI; 2001-639201/73.
XX DR N-PSDB; AAD23451.
XX PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer.

XX
PS
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ
Query Match 3.3%; Score 436; DB 4; Length 4019;
Best Local Similarity 18.3%; Pred. No. 6.5e-15;
Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
QY 34 HTDVGLEQYQHHSRDYASHLSPGSIIP-----QRRRPSLLSEFPQGNERSQELHLPESH 90
Db 513 HSDIGPVT-----DFSSLPQPNVQSSRP--LSEQLDGLSPELDGMVTDGA 559
QY 91 YL-----PELKSEMEFTESKRPLELLPDLRPSPLLATGQAGSBDLTKDRSLTGK 144
Db 560 ILGKLYKIPELGKQVDEL-----FTAVLSPANT----- 588
QY 145 LFPVSPSPPHDTPLELVPPLPSKEELIQNDVRDREITWVEQOIKLKKQOOLEEEA 204
Db 589 -QFTPLPQP-----PPTQLLP-----IHNQDAFSR-MPLMNGLI-----G 632
QY 205 AKPPEPEKPVSPPTESKHSRLVLIIDENRKAEEAHRILEGLGQVPELPLYNQPSDTR 264
Db 623 SSPHLPHSLPPGSLGTSAQAQSSYPDARDKNSAFN-----PMASDFNNS- 669
QY 265 QYHENIKINQAMRKLLILYFKRNHARKOWKQFCORYDQLMLEALEKKYVERIENRRRA 324
Db 670 -----WTSS-----APTVEG-ENDTWSNA 687
QY 325 KESKVEYVEKOPPEIRKQRELQERMQSVRGSGLSMSAARSEHEVSEIIDGLSEQEN 384
Db 688 QRSTLK--WEK----- 696
QY 385 LEKOMRQLAVIPPLYDADQRIKFINMGLMADPMVKYKDRQVMNMWSEKETEKEFK 444
Db 697 -REALGEMATVAPLY-----TWINPNEKEEPD--WTRVQIAKLWK----- 739
QY 445 MOHPKNFGLIASFLERKTVAECVLYYYLTKNENYKSLVRSYRRRGKSGQOQOQOQOQ 504
Db 740 -----ASSQERAP-----YVQKARDN-----RAALRINKVQMSNDSMKRQ 775
QY 505 QOQOQOPMPSRQOEKDEKE 564
Db 776 QQSDIDPSRIDSELFKDFLKQRESEHEQE----- 805
QY 565 ASKGRKTANSQGRKGR--TRSMANEANSEBAITPQSAELASMEELNESSRTEEMET 622
Db 806 -WFKRQOMRQKSKQAKIETATKLEQVKNQEQ-----QOQOQOQSGSHL-----LVQSGSDT 856
QY 623 AKKGLEHGRNWSAIAARMVSKTVSQCKNFYNYKQRNLDLILQOHLKMEKERNARK 692
Db 857 PSSGI-----QSPLTPQPGNGMSPAQSF-----HKELFTKQ----- 888
QY 683 KKKAPAAASEEAAF-----PPV-----VEDEMEASGVSGNEEMVEAEALHASGN 729
Db 889 PFSTPTSTSDSVFKVQPPAPPPPPAPSRIPIDLSQA----- 936
QY 730 EVPRGCSGPATVNNSSDTEIPSPHTEAAKDTGONGKPPATLGADGPPGPPPTPP--- 786
Db 927 -----QTSQPPSPQVFS-----PGSSNSRPPSPMDPYAKMVGTPRPPVPG 966
QY 787 -----RTSRAPIS-TPASEA-----TGAPTPPPAPP-----SPSAPPVVPK 824

Disclosure; Page 309-318; 378pp; English.

The invention relates to isolated lung tumour-specific proteins and their corresponding cDNA molecules. Lung tumour-specific proteins and their antigen-presenting cells are useful for stimulating and/or expanding T cells specific for a tumour protein, and for inhibiting the development of cancer. The invention also relates to a composition useful for stimulating an immune response, and for treating cancer. The lung tumour specific oligonucleotide is useful in gene therapy and for diagnosis, detection and treatment of lung cancer. The present sequence is human lung tumour-specific protein

Sequence 4019 AA;

Db 967 HSFSSRNSAAPVENCPTPLSVSRPLQMNETTANRPSVVRDLCSSTNNNDPVAKPDP- 1025
QY 825 EKEEETAAPVVEEGEOKPAAAEALAVDTGKAEPVKSECTEEAEPAKGDAEAAE 884
Db 1026 -----RPVMTDQPKSLGLSRSPVISEQT-----AKG---PIAA 1056
QY 885 ATAEGALKAEBGGGRATTAKSGAPODSSSATCSADEVDEAGGDKNLLGPRPSL 944
Db 1057 GTSDHFTKE-----SPRADVFQORIP--DSYARPLLTPAPLDSGPGPKTP--MQPPSS 1108
QY 945 LPTGDPANASPOKPLDLKQLKORAAAIPTQVTKVHEPPREDAAATKAPAPPPQN 1004
Db 1109 QDPYGSV--SOAGRLSVD----PYERPALTPRIDNFNSHNQNDPYSQPLTPHFAVNES 1163
QY 1005 LQPEDAPOQPGSSPRGSKR-----SPAPPADKEAFAAEAQKLPGPCCWTSGLPPP 1056
Db 1164 FAHPSRAFSQPGTIISPTSQDPYSQPPCTPRPVVDSYSSSGTARNTDYPQPGTTPP 1223
QY 1057 --VPREVITAKSHAPDPSAFS-----YA--PFGHPLP--LGLHDTARPVLP 1097
Db 1224 TTVDP--YSQOPOTPRSTQTDLFVTPVINOHRHSDPYAHPGTGPRGISVPYSQPPATP 1280
QY 1098 RP-----PTI--SNPPLISSAKH-----PSVLEROIGAISQMSVOLHVPYS 1138
Db 1281 RPRISEGFTRSSNTRVLMENQDPFLQAAQNRGAPALFGLVLRPPDTCQ----- 1329
QY 1139 EHAAPVGPVTWGL-----PLMDPKKLAPSGVKQ-----EOLSPR 1175
Db 1330 --TPRPPGP--GLSDTFSRVSPAARDPYDQSPMTFRSQSDSFGTSTAHADVADQPRG 1384
QY 1176 GQAGPPESLGVPTAQBASVLRGALGSVPGSGITKGPST----- 1215
Db 1385 SEGSCASSNPMHSQOQFSG--VSQLPFGVPTSGVTDONTVNMAQADTEKLROKXL 1442
QY 1216 -----RVPSDS--AITYRGSIHTGTPADV-----LYGNTITRII 1247
Db 1443 REILLOQOQKXIAGRQSGQSDPAVPHGFLQHWQPNVNAQFTRPPPPYFGNIR-- 1499
QY 1248 GEDSPRLDRG--REDSLPLKGH-----VIYEKKGH--VLSYEGQMSVTQCKEDGRSS 1297
Db 1500 ---SPVAPLGPRIYAVFPKQRGYPDPDVASMGWRPHGFRFGFGSGHGTMPQOE---RF 1553
QY 1298 SGPPHT-----AAPKR-----TYDMEGRVGRATISSIEGLMGRATIPRERHSHHLK--EQ 1348
Db 1554 LVPPQOIQSGSVSPQLRRSVSDM-----PRPLNNSQMNVPGL-----PQHFSQSLPVQO 1605
QY 1349 HHTIRGI-----TQIGPRSYVEAQEDYLREAKLLKRGTPPP--PPPS 1390
Db 1606 HNILQOAYITELRHRAPDGRQRLPFSAPPGSVVEASSNL--RHGNFTPRDPFGPRHTDPM 1663
QY 1391 RDLTEAYKTQALGPLKLKPAHEGLVATVKEAGRSIH--BIPRELRHT--PELPLAPRPL 1446
Db 1664 RPPQGLPNQ-----LPVHFDLEQVPSQEQGHSHVSSWMVRLNHPILGGESEAPLST 1719
QY 1447 KEGSITQGTPLKYDTCASTGSKKHVRSILGSPGRTFFPVHPLDMADARALE--RACYE 1505
Db 1720 SVPSETSDNLQITTPQSDGLEKLSDD-----PSVKELDQ--KOLEGVEVKDLDD 1769
QY 1506 ESLKSPGTASSSGGSJAR-----CAPVIV-----PELCKPQSPPL 1541
Db 1770 EDLENJ--NLDTEGKVKVELTDLNETNDPNLDDLLRSGEFDIIAYTDPFLDMGDKKSM 1827
QY 1542 TVEDHGAPFAGHLPRGSPVTMRPTPRLQEGS--LSSSKASQDRKLTST----- 1588
Db 1828 FNEELDLPIDDKL--DNQCVSVPEKKQEENKTLVLSDKHSQKSVTVNEKTEVLSPN 1885
QY 1589 -----PREIAKSHSTVPEHHP-----HPISP--YEHLLR----- 1616
Db 1886 SKVESKETEKENDKNDVTPCSQASASDLNDGDKTSLHPCDPLFEKRTNRETAGPS 1945
QY 1617 -----GVSVDLYRSHIPLA----FQPTSI--PRGIPLDAAAAYLPR 1653
Db 1946 ANVIOASTQLPAQDVINSGCITGSTPVLSSL--LANEKSNSDIRPSGSP----- 1993

QY 1654 HLAHPNTYP-----HL--YPPYLIRGYPTDALENRQTIINDYIT--SOQMHTTATAMAQ 1705
Db 1994 ---PPTLTPASPSNVSSLPPIA---PPGRVLDN---AMNSNVTVSVRNVH---VFQ 2040
QY 1706 RADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPPTPGTPATAMDRLAYLPTAP 1765
Db 2041 GYVQNPGLIPGOST--VNHSLG-----TGKPAT-----QTGP 2070
QY 1766 QPFSRSHSSPLSPGPGTHLTKPTTSSSERER-----DRDRERDRDRER 1810
Db 2071 Q--TSQSGTSSMS--GPOOLMIPQTLAQNRRERPLLEEQLLLQDLDOERQEQOQO 2126
QY 1811 EKSIIITSTT--TVEHAPIWRPGTEQS----- 1834
Db 2127 MQAMIRQRSEPPFNIDFDAITDPIKAKWALKGINKVMAQNNGMPVMVMSRPFPMQ 2186
QY 1835 --SGSGSGSGGGGSSSRPASHAHQHSPISP-----RTQDAL 1871
Db 2187 VVTGTQNSGQNLGPAIPODGSITHQISRPNPNFPGFVNDQSRKQVEEWLQETQQL 2246
QY 1872 QORPSVL-----HNTGMKGIITAVEPSKPTVLRSTSTSSVVRPAATPPATHCPLGGTL 1925
Db 2247 QMOQKYLEEQIGAHRSKKAL-----SAKQRTAKKAGREFFPEDEAOLKHVTE-- 2294
QY 1926 DGVIPTLMEPVLLPKXAPRVARPERPRADTGHAFIAK-----PPA-----RSGLE 1970
Db 2295 -----QOSMVQKQEQIRKQKQEAELIEDYIRKQOQOCAMAPPTMMPVQOPPLJ 2346
QY 1971 PASSPSKSGEPR--PLVPP-----VSGHATTARTPA-----KNLAPHASDDPP-- 2012
Db 2347 PGATPPTMSQPTFPMVPOLOHQOHTTIVISGTSVPRMPSLPGWQFNSAPAHLPNPPRI 2406
QY 2013 -----APPASASDPHEKTSQKPPS----- 2032
Db 2407 QPIIAQLPIKCTPAGTVSNANPQSGPPRVFEDDNNPFSSEFORERKEBLRQOERQ 2466
QY 2033 -----IQELELSRGLYHSGSYSP-----EGVEYSPVSS 2061
Db 2467 RIQLMQEVDQRALQRMEMEQMGVSEISSRSTVSQIPFYSSDLPCDMQPLGLQ 2526
QY 2062 PSLTHDKLGPLKHELE-----DKSHLEGLRP-----KQPGVKLGGEAAHLP 2104
Db 2527 S-----PQOQOQGVLOQOQNIQOQSINSFSTQTFMOTNERRQVGPSPVSPSP 2578
QY 2105 HLRPLPES-----QPSSSPLLQTAGVKGHQVWVTLAQHISEVITQD 2146
Db 2579 VGSNPFSSVKQGHNLGSGTSFOOSPVRFSTPALPAAPV-----ANSSLPCCQD 2628
QY 2147 YTRHHPQOLSAPLAPLYSFPGASCVPDLDRPPDLPLPPDHGAPARGSPHSEGGKRS 2206
Db 2629 STITHG-----HSYPGSTQSLIQLY---SDII-----PEEKKKKR 2661
QY 2207 PRENTKSLGGGEDGIEBVPSPPEGMT---EPGHSRAVYPLLYRDOEQTEPERMSKSPG 2263
Db 2662 TRKKKRD--DDAESTKAPSTPHSDITAPPTPGISETTSTPAVSTPSLPPQADQESVEV 2719
QY 2264 NTSQPPAPFSKL--TESNSAMVKSQKQINKLNTNREPEYNI SQPGTEIFNMPAITCT 2322
Db 2720 GPSTNMAAGQCTELEKLP-----NSDFSQATPNQOQTYANSEVDKLSMETPAKT-- 2770
QY 2323 GLMTRYSAQVQASHASTNMGLEAIIRKALMGKYDQWEEPSPLSANAFNPNLASIASLPAAMP 2382
Db 2771 -----EEIKLEKAETESCPG-----QEEPKLEEQNGSKVEGNA---VACP 2807
QY 2383 ITAADGRSDHUTISP---GGGKAKVSGRPSRKAASPAAGLASGDRPPSVSVSEGC 2439
Db 2808 VSSAQS--PPHSAGAPAAKGDGSGNELLKHLKNKKSSS-----LLNQKPE--GSI 2859
QY 2440 NRTPLTNVRWDRDRSSAGSTFPYFNPLMRLOAGVMASTPPPPGLPAGS----- 2488
Db 2860 TKONKL---VEKQNAEGLQOT-----LGAQMGGFGCGNQLPKTDGSGSETKKQSRKTQ 2910

QY 2489 --CPLAGPH--HAWDEEPKPLCSQVETLS 2514
 Db 2911 RTGEKAAPRSKRRKXDEEKQAMYSSTDTFT 2941

RESULT 57
 ADD66733
 ID ADD66733 standard; protein; 4019 AA.
 AC ADD66733;
 XX 15-JAN-2004 (first entry)
 DT Human lung tumour-specific related protein, SEQ ID NO 425.
 XX expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;
 KW human; lung tumour-specific.
 XX Homo sapiens.
 OS WO200292001-A2.
 PN 21-NOV-2002.
 PD 10-MAY-2002; 2002WO-US014975.
 PF 11-MAY-2001; 2001US-00854133.
 PR (CORI-) CORIXA CORP.
 XX Lodes MJ, Wang T, Fan L, Algate PA, Mcneill PD;
 PI WPI; 2003-120592/11.
 XX New polynucleotide and polypeptide, useful for preparing a composition
 PT for diagnosing, treating or preventing cancer.
 XX Disclosure; SEQ ID NO 425; 494pp; English.

The invention relates to a novel isolated polynucleotide comprising one
 of 32 47-6080 base pair sequences, given in the specification, or their
 complements or degenerate variants, at least 20 contiguous residues of a
 sequence in, or having at least 75 or 90 % identity with the isolated
 polynucleotide, or that hybridise with the polynucleotide. The invention
 further comprises: an isolated polypeptide; an expression vector
 comprising the polynucleotide operably linked to an expression control
 sequence; a host cell transformed or transfected with the expression
 vector; an isolated antibody or its antigen-binding fragment that
 specifically binds to the polypeptide; a method for detecting the
 presence of a cancer in a patient; a fusion protein comprising the
 polypeptide; an oligonucleotide that hybridises to the isolated
 polynucleotide under moderately stringent conditions; a method for
 stimulating and/or expanding T cells specific for a tumour protein; an
 isolated T cell population; a composition comprising a first component
 consisting of carriers and immunostimulants and a second component; a
 method for stimulating an immune response in a patient; a method for
 treating cancer in a patient; a method for determining cancer in a
 patient; a diagnostic kit comprising at least one oligonucleotide or
 antibody and a detection reagent comprising a reporter group; and a
 method for inhibiting the development of cancer in a patient. The
 compositions of the invention have cytostatic activity and can be used to
 create a vaccine. The isolated polynucleotide is useful for preparing a
 composition for diagnosing, treating or preventing cancer. This sequence
 CC represents a human lung tumour-specific protein relating to the
 CC invention.

Sequence 4019 AA;
 SQ Query Match 3.3%; Score 436; DB 7; Length 4019;
 Best Local Similarity 18.3%; Pred. No. 6 5e-15;
 Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
 QY 34 HTDVGLEYYQHSDRYASHLSPGSIITQP---QRRRPSSLSEFQGNERSQELHLRPESH 90

Db 513 HSDIGVTD-----DFSSLPQPNVNOSSRP--LSEBQLDGILSPELDKNMTDGA 559
 QY 91 YL-----PELGKSEMEFIESKRPRLELLPDLPLRSPILLATCQAPAGSEDLTKDRSLTGK 144
 Db 560 ILGKLYKIPELGGKVEDL-----FTAVLSPANT-----588
 QY 145 LEPVSPSPPTDPELVLVPPRLSKKEELTONMDRVDRITMVEQQISKUKKQQQLEBEA 204
 Db 589 -QPTPLPQP--PPTQLLP-----IHNQDAFSR-MPLMNGLI-----G 622
 QY 205 AKPPEPEKVPSPPIESKHSLSLVQIIYDENRKAEEAHRILEGLGQVVELPLYNQPSDTR 264
 Db 623 SSPHLPHNSLPPGSGGLGTFSATQAQSSYPDARDKNSAFN-----PMASDPNNS- 669
 QY 265 QYHENIKINQAMRKKLILYFKERNHARKQWKQFCQRYDQLMEALEKKYVERIENNRRA 324
 Db 670 -----WTSS-----APTVEG-ENDTMSNA 687
 QY 325 KESKRVREYVEKOPPEIRKQRELOERMQSRVGQSGLSMSAARSEHEVSEIIDGLSEQEN 384
 Db 688 QRSTLK--WEK-----696
 QY 385 LEQMRQLAVIPMLYDADQORIKFINMGLMADPMKYVKDRQVMNMSEQEKETREXF 444
 Db 697 -EALGEMATVAPVLY---TNINPENLKEEFPD--WTRVKQIAKLMRK-----739
 QY 445 MQHPKNFGLIASFLERKTVAECLVLYLTCKENYKSLVRRSYRRRGKSGQQQQQQQQQQ 504
 Db 740 -----ASQERAP-----YVOKARDN-----RAALRINKVQMSNDSMKRQQ 775
 QY 505 QQQQQQMPRSSQBEKE 564
 Db 776 QQSDIDPSSRIDSELPKPLKQRESEHEQE-----805
 QY 565 ASKGRKTANSQRRKGRIT--TRSMANEANSEBAITPOQSAELASLMELNSSRTESEMET 622
 Db 806 -WKFRQOMRQSKQAKIEATOKLEOVKNEQQ---QQQQQQQFGSQHL-----LVQSGSDT 856
 QY 623 AKKGLELHGRNNSATARMVSGKTSOCKNFYFNYKKRQNLDELILQHLKLMKERNARRK 682
 Db 857 PSGLI-----QSPLTPQPGNGNMPAQSF-----HKELFTKQ-----888
 QY 683 KKKAPAAASEEAAF-----PPV-----VEDEMEASGVSGNEEMVEEAEALHASGN 729
 Db 889 PSTPTSTSSDDVFVKPQAPPPPPAPSRIPIDQSLSQ-----926
 QY 730 EYPRGECGPATVNNSSDTSIPSPHTEAAKDTGQNGPKPATLGADGPPPGPTTP--786
 Db 927 -----QTSQPPSPQVFS---PGSSNSRPPSPMPDYAKMVGTPRPPVPG 966
 QY 787 ---RRTSRAPTEP-TPASEA-----TGAPTPPPAPP-----SPSAPPVVPVK 824
 Db 967 HFSRNRNSAAPVENCPTLSSVSRPLQWNETTANRPSFVDRCLCSSSTNNNDPYAKPDTP- 1025
 QY 825 EKEEETAAPVVEGEQKPPAAEALAVDTGKAEPPKVEKSECTEEAEAGPKAGKDAEAE 884
 Db 1026 -----RPVMTDQFKSLGLSRSPVVSQI-----AKG---PIAA 1056
 QY 885 ATAGALKAEKKEGGSGRATTAKSGAPQDSSSATCSADEVDEAGGDKNRLLSRPSL 944
 Db 1057 GTSDHFTK-----SPRADVFQRIIP--DSYARPLLTAPLDSGFGPFTKTP-MQPPSS 1108
 QY 945 LPTGDPANASPKPLDLKQLKQRAAAIPIQVTKVHEPPPREDAAPTAPAPPPPPON 1004
 Db 1109 QDPYGSV-SQASRRLSVD-----PYERPALTTPRIDNFSHNQSNNDPYSQPLTPHPAVNS 1163
 QY 1005 LOPESDAPQPGSGSPRGKSR-----SPAPPADKEAFAAEAQKLPGBPCCWTSGLPFP 1056
 Db 1164 FAHPSRAFSQPGTISRPTSDPYSQPPGTPRPFVDSYSQSSGTARSTNTDPYSQPGTTPR 1223
 QY 1057 --VPREVITKASHPADPSAFS-----YA-PPCHPLP-LGLHDTARPVL 1097

Db 1224 TTVDP---YSQQPQTPRPSTQTDLFVTPVTNQRHSDPYAHPPTPRGIVSVFSPQPPATP 1280
Qy 1098 RP-----PTI-SNPPLISSAKH-----PSVLERQIGAISQGMSVOLHVPYS 1138
Db 1281 RPRISGFTSSMTRPVLMPNQDFFLOAQRNGRGPALPGLVPRPPTCSQ----- 1329
Qy 1139 EHAKAIVGPTWGL-----PLPMDPKKLAPFGVKQ-----EQLSPR 1175
Db 1330 --TPRPGP--GLSTDFSRVPSAARDPYDQSPMTPRSQSDSFGTSQTAHDVADQPRPG 1384
Qy 1176 GQAGPESLGVPTAQASVLRGALGVPQSGSTTKGIPST----- 1215
Db 1385 SEGSCASSNPMHSQCGQFSG--VSQLEGPVPTSGVTDTQNTVNMAQADTEKLQRQKL 1442
Qy 1216 -----RVPSDS-AITYRGSITHTGPADV-----LYKGTITRII 1247
Db 1443 REILIOOQOKKIAGROEGSQSPAVPHGPLOHQWOPENVNAQFTRPPPPYFGNIR---- 1499
Qy 1248 GEDSPRLDRG-REDSLPKGH-----VIYEKGKGH--VLSYEGGMSVTQCKEDGRSS 1297
Db 1500 ---SPVAPPLGPYAVFPKQRCGPYPDVASMGMRPHGFRFGPGGSHGTMPQOE---RF 1553
Qy 1298 SGPHET---AAPKR---TYDMGVRGVRATSSAIEGLMGRATPERRHSPHHLK-EQ 1348
Db 1554 LVPPQIQGSGVSPQLRRSVSDM-----PRLLNSQMNPNVGL---POHFSQSLPVQO 1605
Qy 1349 HHIRGSI-----TQGI PRSYVEAQEDYLREAKLLKRGTPPP--PPPS 1390
Db 1606 HNILGOAYIELRHAPDGRQRLPFSAPPQSVBASSNL--RHGNFIPRDPDFGPRHTDPM 1663
Qy 1391 RDLTEAYKTQALGLPKAHEGLVATVKEAGRSIH--BIPRELRHT--PELPLAPREL 1446
Db 1664 RRPQGLPNO---LPVHPDLEQVPPSQEQGHVSHSSWMVETLHNLPLGSEFSEAPLST 1719
Qy 1447 KEGSITQGTPLKYDTGASTGSKKHVRSIGSPGRTFPPVHPDLVMADARALE-RACYE 1505
Db 1720 SVPSETTSDNLQITTPSDGLEKLSDD-----PSVKELD-V-KOLEGVEVVDLDD 1769
Qy 1506 ESLKSRPGTASSGGSIAR-----GAPVIV---PELGKPRQSPL 1541
Db 1770 EDLENL--NLDTEGKVVELDTLDNLETNDPLDLDLRLSGEPDIIAYTPELMDGDKSM 1827
Qy 1542 TYEDHGAFFAGHLPRGSPVMTREBTPRLQEGS--LSSSKASQDKLST----- 1588
Db 1828 FNEELDLPIDDKL--DNQCVSVEPKKEQENKTLVLSDRKSPQKSTVTNEVKTEVLSPN 1885
Qy 1589 -----PREIAKSPHSTVPEHIP-----HPISP--YEHLLR----- 1616
Db 1886 SKVESKETEKNENKDNVDTPCQASAHSDLDNGEKTSILHPCDPLFEKRTNRETAGPS 1945
Qy 1617 -----GVSGVDLYRSHIPLA--FDPTSI-PRGIPLDAAAAYVLP 1653
Db 1946 ANVIQASTQLPAQDVINSCGITGSTVLSL-LANEKSDNSDIRPSGSP----- 1993
Qy 1654 HLAFNPTYP-----HL--YPPYLIRGYPDTAALENRQTIINDYIT-SQQMHNTATAMAQ 1705
Db 1994 ---PPPTLPASPNSHVSLPFFIA---PPGRVLDN--AMNSVTVVSVNH---VFSQ 2040
Qy 1706 RADMLRGLSPRESLAINAAGPRGIIDLQVPHLPVLVPTPTGTATAMDRLAYLPTAP 1765
Db 2041 GVQVNPGLIPQGST--VNHSLG-----TGKPAI-----QTGP 2070
Qy 1766 QPFSSRHSSSPPLSGGTHLTKTPTTSSRRER-----DRDRDRDRDR 1810
Db 2071 Q--TSOGSTSSMS--GQQUMIQTUAAQNRERPLLEEQPLLQDLDLDOEREOQQORQ 2126
Qy 1811 EKSILTSTT-----TVEHAPIMRPGTEQS----- 1834
Db 2127 MQAMIRQSRSEFPFNIDFALITDPIKAKVVALKGINKVMAQNGLMPPMWSRFFPMGQ 2186
Qy 1835 --SGSSGSGSGGSSSRPASHAHQHSPISP-----RTQDAL 1871
Db 2187 VVTGTQNSQGNLGPQAIPODGSITHQISRPNPNFPGFVNDSPQRKQYBEWLQETQLL 2246

Qy 1872 QORPSVL-----HNTGMKGIIITAVEPSKPTVLRSTSTSSSPVRPAATFPFATHCLPGTGL 1925
Db 2247 QWQKYVLEBQICAHRSKKAL-----SAKQRTAKAGREFFEDAEQLKHVTE----- 2294
Qy 1926 DGVTPTLMEPVLLPKAARVARPERPRADTGHAFIAK-----PPA-----RSGLE 1970
Db 2295 -----QOSMVQKQLEQIRKQKEHAELIEDYRIKQOQOCAMAPPTMPSVQOPPLI 2346
Qy 1971 PASSFSKSEPR-PLVPP-----VSGHATTARTPA-----KNLAPHASDPP-- 2012
Db 2347 PGATPTWSPQSTFPMVPOOLQHQHTTVISGHTSVPRMPSLPGWQPNAPAHPLNPPRI 2406
Qy 2013 -----APPASADPHREKTSKPPFS----- 2032
Db 2407 QPIAQLPIKCTPAPGTVSNANPOSPPRPFVEFDDNNPFSFQERERKERLRBQQERQ 2466
Qy 2033 -----TOELELSRGLVHGSSYSP-----EGVEPVSPVS 2061
Db 2467 RIQLMQEVDQRALQORMEMEQHGMVGBEISSRSTSVSIPIFYSSDLPDCFQMLGPLQO 2526
Qy 2062 PSLTHDKGLPKHLEEL-----DKSHLEGELRP-----KQPGVVLGGAAHLP 2104
Db 2527 S-----POHQOQMGVLOQONIQGSSINSPTQTFTWTRERQVGPSPFVDPSPSTP 2578
Qy 2105 HLRPLPES-----QFSSSPLLQOTAPCGVKGHRVVTILAOHISEVITQD 2146
Db 2579 VCSNPFSSVKQGHGNSLGTSTFQOSPVRPFTALPAAPV-----ANSSLPCGD 2628
Qy 2147 YTRHHPQQLSAPLAPLPSVFPFGASCPLVDLRPPSDLYLPPDPHGAPARGSPHSEGGKRS 2206
Db 2629 STITHG-----HSYPGSTQSLIOLY---SDII-----PEEKGKKR 2661
Qy 2207 PEPNKTSLVGGEDGIEPVSPPEGMT---EPGHSRSVAVVPLLYRDGEQTEPSRMGSKSPG 2263
Db 2662 TRKKRDR--DDAESTKAPSTPHSDITAPPTPGISETTSTPAVSTPSLPPQADQESVEPV 2719
Qy 2264 NTSQPPAPFSKL--TESNSAMVKSKEQINKLNTNRNEPEYNI SQPGTEIFNMFAITGT 2322
Db 2720 GFSTPNMAAGLCTELENKLP-----NSDFSQATPNQOTYANSEVDKLSMETPAKT-- 2770
Qy 2323 GLMTRYSQAVQEHASHTNMGLEAIIRKALMGKYDQWEEESPPLSANAFNPLNASLPAAMP 2382
Db 2771 -----BEIKLEAETESCFG-----QEEPKLEEONGSKVEGNA---VACP 2807
Qy 2383 ITAADGRSDHTLTSP---GGGKAKVSGRPSRKAAPGLASGRDRPPSVSVHSEGD 2439
Db 2808 VSSAQG--PPHSAGAPAAKGDGSGNELLKHLKNKKSS-----LLNQKPE--GSI CSEDDC 2859
Qy 2440 NRTPLTNRVWEDRPSAGSTPFPYNPLIMRIQAGVMASSPPPPGLPAGS----- 2488
Db 2860 TKDNKL--VEKQNPAGELQI-----LGAQMGGFGCGNQLPKTDGSETKKQSRKTQ 2910
Qy 2489 --GPLAGPH--HAMDEEPKPLLCQYETLS 2514
Db 2911 RTGEKAAPRSKKRKKDEEEKQAMYSTDTFT 2941

RESULT 58

AD8E87987
ID ADE87987 standard; protein; 4019 AA.
XX
AC ADE87987;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human lung tumour antigen polypeptide #92.
XX
KW Human; lung tumour antigen; cancer; lung cancer; CD8+; T cell;
KW immune response; immunostimulant; cytostatic.
XX
OS Homo sapiens.
XX

PN US2003118599-A1.
XX 26-JUN-2003.
XX 10-MAY-2002; 2002US-00144649.
XX 02-APR-1999; 99US-00285323.
XX 09-AUG-1999; 99US-00370838.
XX 30-DEC-1999; 99US-00476235.
XX 03-MAR-2000; 2000US-00518809.
XX 29-MAR-2000; 2000US-00538037.
XX 05-JUN-2000; 2000US-00588937.
XX 18-AUG-2000; 2000US-00640878.
XX 20-SEP-2000; 2000US-00667170.
XX 01-NOV-2000; 2000US-00704512.
XX 14-DEC-2000; 2000US-00738973.
XX 11-MAY-2001; 2001US-00854133.
XX (CORI-) CORIXA CORP.
PA Algate PA, Lodes MJ, Wang T, Fan L, Mcneill PD;
PI WPI; 2003-897103/82.
DR N-PSDB; ADE87984.
XX
PT New polynucleotides encode lung tumor antigens and are useful to
PT stimulate an immune response or detect or treat a cancer in a patient,
PT particularly lung cancer.
XX
PS Disclosure; SEQ ID NO 425; 63pp; English.
XX
CC The invention relates to polynucleotides encoding lung tumor antigens.
CC The invention also relates to the polypeptides encoded by the
CC polynucleotides, isolated antibodies or antigen-binding fragments that
CC specifically bind the polypeptides and a method for detecting cancer in a
CC patient, comprising obtaining a biological sample from the patient,
CC contacting the sample with a binding agent that binds a polypeptide of
CC the invention, detecting in the sample an amount of polypeptide that
CC binds to the binding agent, and comparing the amount of polypeptide to a
CC predetermined cut-off value. T cells specific for a tumor protein can be
CC stimulated and/or expanded by contacting the T cells with a polypeptide,
CC polynucleotide or an antigen-presenting cell that expresses a
CC polypeptide. Cancer development can be inhibited by incubating CD4+
CC and/or CD8+ T cells isolated from a patient with a polypeptide,
CC polynucleotide or an antigen-presenting cell that expresses a
CC polypeptide, so that the T cells proliferate. The invention is used to
CC stimulate an immune response or to detect or treat a cancer in a patient,
CC particularly lung cancer. This sequence represents a human lung tumor
CC antigen polypeptide of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification but was obtained in
CC electronic format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 4019 AA;
Query Match 3.3%; Score 436; DB 7; Length 4019;
Best Local Similarity 18.3%; Pred. No. 6.5e-15;
Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
34 HTDVGLLEVOHHRDYASHLSPGSIQTP---QRRRPSLLSEFOPGNERQELHLRPSHS 90
513 HSDIGFVTD-----DPSLLPQPNVNSRSP--LSEQLDGLTSPDLKMTDGA 559
91 YL-----PELGKSEMEFTESKEPRLELLPDLLRPSPLLATQAPGSEDLTKDRSLTGK 144
560 ILGKLYKIPELGGKVEDL-----FTAVLSPANT----- 588
145 LEVPSPPSPHTDPELELPPRLSKBELIONMDRVDREITWVEQQLSKLKKQOQLEEEA 204
589 -QTPPLPQPP---PPTQLLP-----INQDAFSR-MPLMNGLI-----G 622
205 AKPPEPEKVPSPPIESKHSRLVQIITYDENRKAEEAAHRLIEGLGPQVELPLYNQPSDTR 264
623 SSPHLPHNSLPPGSLGTGTSIAQSSYPDARDKNSAFN-----PMASDPNNS- 669

QY 265 QYHENIKINQAMRKKLILYFKRRNHARKQWKQKFCORYDQLMEALEKKYVERIENNRRRA 324
DB 670 -----WTSS-----APTVEG-ENDTMSNA 687
QY 325 KESKUREYYEKOPPEIRKORELOERMQSRVQGRGSLSMSAARSEHEVSEIIDLGEQEN 384
DB 688 QRSTLK--WEK----- 696
QY 385 LEKOMROLAVIPMLYDADQRIKFINMGLMADPMKVYKDRVMMWMSQEKETPREKF 444
DB 697 -EALGEMATVAPVLY---TWINFNLKEEPPD--WTRVKQIAKLWRK----- 739
QY 445 MQHPKNFGLIASFLERKTVACVLYYYLTTKENYKSLVRRSYRRRKGSKQOQOQOQOQ 504
DB 740 -----ASSQERAP-----YVQKARDN-----RAALRINKVQMSNDKMRQ 775
QY 505 QOQOQOQMPRPSQOEKDEKEKEKEKEKEKEKEKEKEKEDELLKEKTDGTSGEINDEKAV 564
DB 776 QODSIDPSSRIDSELFKPLKORESEHEQ----- 805
QY 565 ASKGRKTANSQGRKGRGRI--TRSMANEANSEEAITPQOSAELASLMELNESSRTEBEMET 622
DB 806 -WFKRQMRQKSKQAKIEATQKLEQVKNQEQ--QOQOQOQFGSQHL-----LVQSGSDT 856
QY 623 AKKGLLIEHGRNWSAIARWVGSKTVSQCKNFYFNKKRQNLDEILQOHLKMKERNARRK 682
DB 857 PSSGI-----QSPLTPQFGNGMSPAQSF-----HKELFTKQ----- 888
QY 683 KKKAPAAASEEAAAF-----PPV-----VEDEMEASGVSGNEBEMVEEAALHASGN 729
DB 889 PSTPTSTSSDDVFVKPQAPPPPPAPSPRIPQDLSLQA----- 926
QY 730 EYPRGCSGPAIVNNSSDTESTPSHTEAAKOTGQNGKPPATLGADGPPPGPTTP-- 786
DB 927 -----QTSQPPSPQVFS--PGSSNSRPPSPMDPYAKMVGTPRPPVPG 966
QY 787 ---RRTSRAPIEP-TPASEA-----TGAPTPPPAPP-----SPSAPPVPVVK 824
DB 967 HSFSRNSAAPVENCFTPLSVSRPLQMNETTANRPSFVRDLCSSTTNDPYAKPDTTP- 1025
QY 825 EKEEETAAAPVVEEGEOKPAAAEAVDTGKAEFPVKSECTEEAEGBGAKOKDAEAE 884
DB 1026 -----RPVMTDQFKSLGLSRSPVVEQT-----AKG-----PIAA 1056
QY 885 ATAEGALKAEKEGGSGRATTAKSSGAPODSSSATCSADEVDEAGGDKNRLLSRPSL 944
DB 1057 GTSDFHTKP-----SPRADVFQORQIP-DSYARPLLTAPLDSGGPGFKTP-MQPPPS 1108
QY 945 LPTGDPANAGPQKPLDLKQLKQRAAAIPIQVTKVHEPPREDAAPTAPAPPAPPPQN 1004
DB 1109 QDPYGSV-SQASRRRLSVD---PYERPALTTPRIDNFHNSQNSNDPYSPPLTTPHAPVNES 1163
QY 1005 LOPESDAPQOQSGSPRGKSR-----SPAPPADKEFAAAEQKLPQDPPCWTGSLPFP 1056
DB 1164 FAHPSRAFSQPGTISRPTSQDYPYQPPGTPRPVDSYSQSSGTARGNTDYPYQPGTTP 1223
QY 1057 --VPPREVTKASPHADPPSAFS-----YA-PPGHPLP-LGLHDTARVLP 1097
DB 1224 TTVDP---YSQQFQTPRPTQTDLFTVPTVNRHSDPYAHPPGTPRGISVPSQPPATP 1280
QY 1098 RP-----PTI--SNPPPLISSAKH-----PSVLERQIGAISQMSVOLHPVYS 1138
DB 1281 RPRISGFTSRSMTRVLMNPNDPFLQAAQNRGPPALPGPLVRPPDTCQ----- 1329
QY 1139 EHAKAPVGVMTGL-----FLMPDKPKLAPFSGVKQ-----EQLSPR 1175
DB 1330 --TPRPPGP---GLSDTFSRVSPSAARDPYDQSPMTPRSQSDSFGTSQTAHDVADQPRG 1384
QY 1176 GQAGPPESLGVPTAQAEASVLRGTALGSVFGGSITKGIPT----- 1215
DB 1385 SEGSGFCASSNPMHSQOQFGSG--VSQLPGPVPTSGVTDTONTVMAQADTEKLQRQKL 1442

PA (HYSE-) HYSEQ INC.
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
DR N-PSDB; AB211953.
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.
XX Claim 9; SEQ ID NO 1783; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (1) comprising a
CC nucleotide sequence selected from any of 948 sequences (AB21119-
CC AB212066) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 4025 AA;
SQ
Query Match 3.3%; Score 436; DB 5; Length 4025;
Best Local Similarity 18.3%; Pred. No. 6.6e-15;
Matches 555; Conservative 314; Mismatches 1010; Indels 1152; Gaps 142;
59 HSDIGPVD-----DPSLLQPNVQSSRP--LSEQLDGLPDLKMTDGA 565
34 HTDVGLEQHSRDYASHLSPGSI1QP---QRRRPSLLSEFOPGNERSGELHLPESH 90
519 HSDIGPVD-----DPSLLQPNVQSSRP--LSEQLDGLPDLKMTDGA 565
91 YL-----PELGKSEFIESKRPRLLELLPDLPLRPSPLATQAPAGSEDLTKDRSLTGK 144
566 ILGLKYIPELGKQVEDL-----FTAVLSPANT----- 594
145 LEPVSPSPPHDPELELPPRLSKEELTONMDRVDRITWVEQIISKLKQQQLEEEA 204
595 -QPTPLPQPP---PPTQLLP-----IHNODAFSR-MPLMNGLI-----G 628
205 AKPPEPEKVPSPPIESKHSRLVQIYDENRKAEEAHLLEGLPQVELPLYNQPSDTR 264
629 SPSPLPNSLPPOSGLGTSAIAQSSYPDARDKNSAFN-----PMASDPNNS- 675
265 QYHENIKINQAMRKLILYFKRNRHARKQWKQKFCQYDQLEALEKKYERIEENPRRA 324
676 -----ASSQERAP-----YVQKARDN-----APTVEG-ENDTMSNA 693
325 KESKVBREYKQPEIRKQBELQERMQSRVQGRGSLMSAARSEHEVSEIIDGLSEQN 384
694 QRSTLK--WEK----- 702
385 LEKOMQLAVIPMLYDADQRIKFINMGLMADPMKYVKDROVMNWSGEKETFREKF 444
703 -EALGEMATVAPLVY-----TNINFNLKEEFPD--WTRVKQIAKLWRK----- 745
445 MQHPKNGFLIASFLERKTVAECVLYLTIKNENYKSLVRSYRRRGKSKQQQQQQQQQ 504
746 -----ASSQERAP-----YVQKARDN-----AALRINKVQMSNDSMKRQQ 761
505 QQQQQQPMPSRSSQEEKDEKEKEKEBEKPEVENDKEDLLKEKTDGSDGDNDEKEAV 564

782 QQSDIDPSRIDSELFKDLKQRESEHQE----- 811
565 ASKGRKRTANSQGRKRGRI--TRSMANEANSEBAITPQQAELASMELNSSRWTEEMET 622
812 -NKFRQMRQKSKQAKIEATQKLEQVNEQQ--QQQQQQFGSQHL-----LVQSGSDT 862
623 AKKGLLEHGRNWSATARMVGSKTVCQCNFYNYKGRQNLDELLOQHKUKMEKERNARRK 682
863 PSSGI-----QSPLTPQPGNGNMSPAQSF-----HKELFTKQ----- 894
683 KKKAPAAAESEAAAF-----PPV-----VDEEMEASGVSGNEEMVEEAEALHASGN 729
895 PSTPTSTSSDDVFKVQAPPPPPAPSRPIQDLSQA----- 932
730 EVPRGECGPATVNNSSDTEIPSPHTEAAKDTQNGPKPPATLGADGPPPGPPTTP-- 786
933 -----QTSQPPSPQVFS-----PGSSNSRPPSPMDPYAKMVGTPRPPVG 972
787 -----RTSRAPIEP-TPASEA-----TGAPTPPPAPP-----SPSAPPVPVVK 824
973 HSFSRRNSAAPVENCCTPLSSVRPLQMNETTANRPSFVRDLCCSSSTNNNDPYAKPPDTP- 1031
825 EEKEEBETAAAPVEEGEEOKPPAAABELAVDTGKAEBPVKSECTEEAEEGPAKCKDAEAE 884
1032 -----RPVMTDQPKSLGLSRSPVSEQT-----AKG-----PIAA 1062
885 ATAEGALKAEEKGSGRATTAKSCAPQDSSSATCSADEVDEARGDKNRLLSRPSL 944
1063 GTSDHFTKP-----SPRADVQRQRIIP-DSYARLLTPAPLDSGPGFKTP-MQPPSS 1114
945 LPTGDDPRANAFQKPLDKQLKQRAAIPPTQVTKVHEPPREDAAPTAPAPPAPPPQN 1004
1115 QDPYGSV-SQASRRLSVD---PYERPALTPTPRIDNFNSQNSNDPYSQPLTPHPAVNES 1169
1005 LQPEDAPQPGSSPRGKSR-----SPAPADKEAFAEAQKLPGDPPCWTSGLPFP 1056
1170 FHPSRAFSQGTISAPTSQDPYSPGPTPRPVVDSYSSGSTARNTDYPSPGTPRP 1229
1057 --VPREVITKASPHADPPSAFS-----YA-PPGHPLP-LGLHDTARPLVP 1097
1230 TVVDP---YSQGPOTPRPSTQTDLFVTVTNQRHSDPYAHPPTPRPGISVPSQPPATP 1286
1098 RP-----PTI-SNPPPLISSAKH-----PSVLEIQIGAISQMSVOLHVPS 1138
1287 RPRISGEFTRSMTRPVLMFNQDPPFLQAAQNRGAPALPGLVRRPDPDTCQ----- 1335
1139 EHAAPVGPVTWCL-----PLMDPKKLAPSGVKQ-----EOLSPR 1175
1336 --TPRPPGP---GLSDTFSRVSPSAARDPYDQSPMTPRSQSDSFGTSQTAHDVADQPRPG 1390
1176 GQAGPPESLGVPTAOEASVLRGTALGSPVGGSIITKIGIPST----- 1215
1391 SEGSCASSNSPMHSGQQQFSG--VSQLPGPVPTSGVTDTONTVMAQADTEKLQRQKL 1448
1216 -----RVPSDS-AITYRGSITHTGTPADV-----LYGTTTRII 1247
1449 REIILQQQKKIAGKQEGSDSPAHPHGPLOHQPENVNQAFTRPPPPYPGNIR--- 1505
1248 GEDSPSLDRG-REDSLPKGH-----VIYEGKKGH--VLSYEGGMSVTCQSKEDGRSS 1297
1506 ---SPVAPPLGPRYAVFPKQDQGPYPPDVASMGMRPHGFRFPFGGSHGTMPSQE---RF 1559
1298 SGPPHET---AAPKR---TYDMMEGRVGRAISSASIEGLMGRAIPPERHSHPHLK-EQ 1348
1560 LVPPQIQSGSVSPQLRRSVSDM-----PRPLANSQMNPNVCL---PQHFSQSLPVQ 1611
1349 HHIRGSI-----TQGISYVEAQEDYLRRKAKLKRGTGTPPP---PPPS 1390
1612 HNILQAVYIELRHAPDGRQLRPFSAAPGVSVEASSNL--RHGNFIPRPDPFGPRHTDM 1669
1391 RDLTEAYKTOALGPLKKAHEGLVATVKEAGSIH--EIPREELRHT--PELPLAPRL 1446
1670 RRPPOGLPNQ-----LPVHPDLEQVPPSQEQGHSHVSSKVMKTLNHLPLGGFSEAPLST 1725

QY 1447 KEGSITQGTPLKYDTGASTTGSKKHVRSLIGSPGRTPFPVHPVLDVMDARALE-RACVE 1505
Db 1726 SVPSSETSNLQITTPSGLEBKLDSD-PSVKELDV-KDLEGVEVKOLDD 1775
QY 1506 ESLKSRPGTASSGGSIAR-GAPVIV-PELGKPRQSP 1541
Db 1776 EDLENL-NLDTEDGKVVLELTDNLNLTNDPLDLRLSGEFDIIAYTDPFLMDGDKSM 1833
QY 1542 TYEDHGAAPAGHLPRGSPVTMRPRLQEGS-LSSKASQDRKLTST- 1588
Db 1834 FNEELDPIDDKL--DNQCVSVPEKKKEQENKTLVLSKHSQKSTVTNEVKTEVLSPN 1891
QY 1589 -PREIAKSPHSTVPEHP--HPISP-YEHLR- 1616
Db 1892 SKVESKETEKENDKNDVTPCSQASAHSLDNGEKTSLHPCDPLFKRTNRETAGPS 1951
QY 1617 -GVSQVDLYRSHIPLA-FDPTSI-PRGIFLDAAAAYLPR 1653
Db 1952 ANVIQASTQLPAQDVINSGITGSTVPLSSL-LANEKSDNSDIRPGSP- 1999
QY 1654 HLAENPTYP-HL--YPPYLIRGYPDTAALENRQTIINDYIT-SQOMHNTATAMAO 1705
Db 2000 -PPPTLPASPNSHVSSLPPFIA--PPGRVLDN--AMNSNVTVSVRNH--VFSQ 2046
QY 1706 RADMLRGLSPRESSLALNVAAGPRGIIDLSQVPHLPVLVPTPTGTPATAMDRLAYLPTAP 1765
Db 2047 GUVQNFGLIPQST--VNHSLG-----TGKPAI- 2076
QY 1766 QPFSSRHSSPLSPGPGTHLTKPTTSSSER-DRDRDRDR 1810
Db 2077 Q--TSQSGTSSMS--GPQQLMIPTQLAQNRRERPLLEQLLQDLDDQERQEQOQ 2132
QY 1811 EKSILSTT--TVEHAPIWEPGTEQS----- 1834
Db 2133 MQAMIRQSEFPFENDFDAITDPIMKAKMVALKINKVMAQNLMGPMWMSRFFPMGQ 2192
QY 1835 --SGSGSGSGGSSSRPASHAHQHSPI-RTQDAL 1871
Db 2193 VVTGTQNSEQNLGPQAIPODGSITHTQISRPNPNFCGFVNDQKQYEWLQETQQLL 2252
QY 1872 QORPSVL-----HNTGMKGIIITAVBSKPTVLRTSTSPVRPAATFPFATHCPLGGTL 1925
Db 2253 QMOQKYLEEQIGAHRSKKAL-----SAKORTAKAGREPPEDAEQLKHVTE- 2300
QY 1926 DGVPYTLMEPVLLPKEAPRVARPERPADTGHAFKAK-----PPA-----BSGLE 1970
Db 2301 -----QOSMVOKOLEQIRKQKQKHAELIEDYRIKQOQOCAMAPPTMPSVQOPPLI 2352
QY 1971 PASSPSKGSER-PLVPP-----VSGHATIAATPA-----KNLAPHASPDPP- 2012
Db 2353 PGATPTMSQPTFMVQQLHQHQTHTVIGHTSPVRMPSLPCGMQNSAPAHPLNPPRI 2412
QY 2013 -----APPASADPHREKTSQKPS----- 2032
Db 2413 QPPTAQLPIKTCTPAPGTVSNANPQSGPPPRVFEDDNNPFESFQERKERLEQOERQ 2472
QY 2033 -----IQELELSRSGYHGSYSP-----EGVFPVSPVSS 2061
Db 2473 RIQLMQEVDQRALQORWEMEGHGWGSEISSRTSVSQIPFYSSDLPCDFMQLPGLQ 2532
QY 2062 PSLTHDKGLPKHLEL-----DKSHLEGLRP-----KQGPVKLOGEAAHLP 2104
Db 2533 S-----PQHQOQMGVQLQQNIQSGINSFSTQFTMTNERRQVGPSPFVDPSPISIP 2584
QY 2105 HLRPLPES-----OPSSPLLOTAPGVKQHVTVLAQHTSEVITOP 2146
Db 2585 VGSFNSFVKQHGNLSGTSFQOSPVRPSPFTPALPAAPPV-----ANSLSLPCQGD 2634
QY 2147 YTRHHPQOLSAPLAPLYSPGASCPVLDLRRPSPDLYLPPDHGAPARGSPHSEGGKRS 2206
Db 2635 STITHG-----HSYFGSTQSLIQLY-----SDII-----PEEKGKKR 2667

QY 2207 PEPNKTSLVGGEDGIEPVSPPEGMT---EPGHSRSVAVPLLYRDGEOTEPSPMGSKSPG 2263
Db 2668 TRKKRD--DDAESTKAPSTPHSDITAPPTPGISSETTSTPAVSTPSELPOQADQESVEPV 2725
QY 2264 NTSQPPAFPSKL-TESNSAMVSKKQEIKNKLTHNRNEPEVNIQSQGTETFNMPAITGT 2322
Db 2726 GFSTENMAAGQLCTELEKLP-----NSDFSQATPNQOQTVANSEVDKLSNETPAKT-- 2776
QY 2323 GLMVTYRSQAVQEHASTNMGLEAIRKALMGKYDOWEESPPLSANAFNPLNASASLPAAMP 2382
Db 2777 -----EEIKLEAETESCPG-----QEEPKLEEQNGSKVEGNA---VACP 2813
QY 2383 ITAADGRSDHTLUTSP---GGGKAKVSGRPSRKAAPGLASGDRPPSVSVHSEGDC 2439
Db 2814 VSSAQSG-PPHSAGAPAAKGDGSGNELLKLLKNKSSS-----LINQKPE--GSIKSEDDC 2865
QY 2440 NRTPLTWRVWEDRSPSSAGSTFPFYNPLIMRLOAGVMASSPPPPGLPAGS----- 2488
Db 2866 TKDNKL---VEKQNPAGELQI-----LGAQMGGFGCGNQLPFTDGGSETKKQSKRTQ 2916
QY 2489 --GPLAGPH---HAWDEEPKPLLCQYETLS 2514
Db 2917 RTGEKAAPRSKKKDEEKQAMYSTDTFT 2947
RESULT 60
ABB60291
ID ABB60291 standard; protein; 2951 AA.
XX AC ABB60291;
DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 7665.
XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL04394.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX PS Disclosure; SEQ ID NO 7665; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at fcp.wipo.int/pub/published_pct_sequences


```
QY 1900 STSTSPVRPAATFFPETHCPGLGTLGVYPTLMEVLLPKAPRVAR-----PE 1949
Db 2073 -TQKETPLPKKQ-----PLSKVKD-----EPEKVNKEPKVPKESQTKLKEEPE 2117
QY 1950 RPRADTGHAFKAPPAR-SGLEPASPSK--GSEPRPLV-----PPV 1988
Db 2118 RVTYKTPQKEPRKEPLRQSEDEPEFSEFDEPLFMTKTHTTAEMKRQKDILNRPSV 2177
QY 1989 SGHATIAI-----TPAKNL-----APHHA 2007
Db 2178 FQORTPERKSTTPSPTKLNGTRGSPSTNLITEKRSYRQVTVNSKPGTRKTPSAN 2237
QY 2008 SPDPAPPASADPHR-EKTSKPSIQ--ELELSRLGVHGSY---SPEGVEPVSPVSS 2061
Db 2238 SPAQSPPTTTSIKRMEQISQSQSVVQDVVDVEVGPAPPESHSEKPGKSP-SPTS 2296
QY 2062 PSIT-----HDKG 2069
Db 2297 RSLRSRSPSPKSIITTYTNTGRNVASRRNVFVHETHVDSEPTGRRRPSYMDHTKS 2356
QY 2070 LPKHLE---ELDKSHL-----EGELRPKQPG-----PVKLGGAAHLPHLRPLP 2110
Db 2357 SLEHTRDSLEINKSHYSKSMEDDSPVEPRNPSSVKFDPVRKSSRGADPEPKTSLK 2416
QY 2111 ESQPSG-----SPLLTQAPGVKGHQVVTLAQHISE-----VI 2143
Db 2417 GKDESDLELETEEIFDLQRLKLETVASYEMRRIRAOMLIRKMNINAGTTTTITI 2476
QY 2144 TDYTRHHFQQLSAPLAPLYFPFGASCVPDLDRPPS-----2181
Db 2477 TTSTT-----PGKSPPLKIRRDQSPAGAAEVKTKVTRTTTERRQQO 2518
QY 2182 -----DLYLPPPDHGAPARGSPHSGGKSPPEPKNTSVLGGEDGIEPVSPPEGMTEPG 2235
Db 2519 QRVEQVDSITPI-----APKTSFPH-----GKPPVKPRERSASPAK---RHSIP-GKQSPG 2567
QY 2236 H-----SRSAVPLLYRD-----GEQTEPSRMGS-KSPGNTSQPP 2269
Db 2568 DRSTTTTTKVTTTSTRGAPSPKPAQGP-IWADRSKVLKGHAIVPQTNGSTPRKSGTSST 2626
QY 2270 AFPSKLTESNAVSKQKQEKLNTHNRNPEYNIQPGTEIFNMPAITGTLMTYRS 2329
Db 2627 SSGKIRTMTS---SSTTTSSSTTNRNKRQREDSI-----TSSYGVGPTDENGULPFGI 2680
QY 2330 QAVQEH-----STNMGLEALIRKA 2349
Db 2681 RALKKATPAEPCETKQVTVGVIEEQFYSDNKSPPRHRKELIYSSNADELAIAKQ 2740
QY 2350 LMGKYDQWESPP-----SANAPNPLNASULP 2378
Db 2741 LQ---DEDDSPPLLDARVVREPKVESQQSLP 2770

RESULT 61
AAE21713
ID AAE21713 standard; protein; 2429 AA.
AC AAE21713;
XX AAE21713;
XX 16-JUL-2002 (first entry)
XX Human PKIN-8 protein.
XX Human; kinase; enzyme; PKIN-8 protein; immune system disorder; anaemia;
KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;
KW asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
KW AIDS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; developmental disorder;
KW Down's syndrome; gene therapy; protein therapy; cytostatic.
XX Homo sapiens.
XX Key
FH Location/Qualifiers
```

```
FT Domain 374. .522 /note= "protein kinase domain"
FT Domain 375. .522 /note= "protein kinase domain"
FT Domain 376. .649 /note= "Eukaryotic protein kinase domain"
FT Domain 378. .636 /note= "protein kinase domain"
FT Domain 379. .522 /note= "protein kinase domain"
FT Domain 489. .507 /note= "Tyrosine kinase catalytic domain"
FT Domain 549. .697 /note= "protein kinase domain"
FT Domain 570. .592 /note= "Tyrosine kinase catalytic domain"
FT Domain 946. .1034 /note= "PDZ domain"
XX WO200218557-A2.
XX 07-MAR-2002.
XX 31-AUG-2001; 2001WO-US027219.
XX 31-AUG-2000; 2000US-0229873P.
XX 08-SEP-2000; 2000US-0231357P.
XX 14-SEP-2000; 2000US-0232654P.
XX 22-SEP-2000; 2000US-0234902P.
XX 29-SEP-2000; 2000US-0238499P.
XX 06-OCT-2000; 2000US-0238389P.
XX 13-OCT-2000; 2000US-0240542P.
XX (INCY-) INCYTE GENOMICS INC.
XX Bandman O, Nguyen DB, Walia NK, Hafalia AJA, Yao MG, Gandhi AR;
XX Gururajan R, Ding L, Patterson C, Yue H, Baughn MR, Tribouley CM;
XX Thornton M, Elliott VS, Lu Y, Ison CH, Au-Young J, Tang YT;
XX Azimzai Y, Burrill JD, Marcus GA, Zingler KA, Lu DAM, Lal PG;
XX Ramkumar J, Warren BA, Kearney L, Policky JL, Thangavelu K;
XX Burford N;
XX WPI; 2002-329769/36.
XX N-PSDB; AAD34305.
XX New human kinases, useful for diagnosing, treating or preventing immune
XX system disorders (e.g. Crohn's disease), neurological disorders (e.g.
XX epilepsy), or cell proliferative disorders (e.g. cancers such as leukemia
XX or lymphoma).
XX Claim 63; Page 156-162; 218pp; English.
XX The present invention relates to human kinases (PKIN) and polynucleotides
XX encoding such proteins. PKIN sequences of the invention are useful for
XX diagnosing, treating or preventing disorders associated with aberrant
XX expression of PKIN, particularly immune system disorders (e.g. acquired
XX immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease,
XX anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-
XX Tooth disease or seizures), cell proliferative disorders (e.g. cancers
XX such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma),
XX and developmental disorders (e.g. Down's syndrome). They are also used in
XX gene therapy and protein therapy. The present sequence is human PKIN-8
XX protein
XX Sequence 2429 AA;
XX Query Match 3.3%; Score 429.5; DB 5; Length 2429;
XX Best Local Similarity 19.9%; Pred. No. 7.8e-15;
XX Matches 551; Conservative 305; Mismatches 982; Indels 931; Gaps 139;
QY 38 GLLEYOH-----SRDYASHLSPGSI-----IQPORRREPSLLSEFQPGNERSQELHLRPE 87
Db 215 GVLSFTTHQIIEIARDCLDKSHQGLTSRYFLELQHLKDLKLL-----QEADRSE 264
```


Db	1973	PSAEPSSSPDPPKPKVAHSS-----ESSSHKPRPGPDGPPKTKHPRDSUSSQKP-	2023
Qy	2032	SI-----QELRLSLG---YHGSSYSPEG--VEPVSPVS-----SPSLTHDKGLPKHLEEL	2077
Db	2024	SVGATKGKEPATQSLGSSREGKHGSKSGPDVFPATPGSONKASDGGICQCGEGGSPVPLHT	2083
Qy	2078	DKSHLEGEIRPKQPG--PVKLGGGAHHPLRLPLPESQSSPSPLLQTPAGVKGHQRVVTLA	2136
Db	2084	DRAPLDAKQPTSGGRPLEVLEKPVHLPRPHGPGSEPADQKL--SAVGEK-----	2132
Qy	2137	QHISEVITQDYTRHHPPQOLSAPLPAPLYSPFGASCVPVLDLR---RPPSDLYLPPPDHGAP	2193
Db	2133	-----QTLSPKHPK-----PSTVKDCP--TLCKQTDNRQTDKSFQ-----P	2167
Qy	2194	ARGSPHSEGGKRSPE-----PNKTSVLGGEDGIEPVSPPEGMTEPHGHSASVYPLLYRDG	2249
Db	2168	AANTDRRAEGKKCTEALYAPAEQDKLEAGLSFVHSENRLKGAERPAAGVGKGF-----	2221
Qy	2250	EQTEPSRMGSKPGNTSQPPAFPSKLTESNAMSVMKSKQEIINKKLNTNRNEPEYNISQP	2309
Db	2222	-----EARGKPGG--PKPP-----TEAD-----KPNGMKRSPP-----	2247
Qy	2310	GTEIFNPAITGTGLMTYRSQAQEH-----ASTNMGLEAIIRKALMGKYDQWEEESPPLSA	2365
Db	2248	-----SATQSSFRSTALPEKSLSCSSSFPETRAGVREASAASD-----	2287
Qy	2366	NAFNPLNASASLPAAPIYAADGRSDHTLTSPGGGK---AKVSGRPSR-----	2412
Db	2288	-----TSSAKAAGMLELPAPSNDRKRAQPAGEGRTHMTKSDSLPSFRVSTLPLESHH	2341
Qy	2413	-----KAKSPAPGLASGDRPPSVSSHSEGDGNCNRRTPLTN	2447
Db	2342	PDPTMGGASHRDRLSVTATVGETKGPAPA-----QPPARKQNVGRDVTKPSAPN	2396
Qy	2448	RWEDRPSS 2456	
Db	2397	---TDRPIS 2402	
RESULT 62			
AAB12000			
ID	AAB12000 standard; protein; 2091 AA.		
XX			
AC	AAB12000;		
XX			
DT	19-DEC-2000 (first entry)		
XX			
DE	Rat p3103 protein.		
XX			
KW	Rat; p3103 protein; Synanon; SH3 domain; PDZ domain; SAPA1 interaction;		
KW	neuronal function; nervous system; neurological disorder.		
OS	Rattus sp.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	554..613	
FT		/note="SH3 domain"	
FT	Domain	654..749	
FT		/note="PDZ domain"	
XX	JP2000184884-A.		
XX			
PD	04-JUL-2000.		
XX			
PF	16-NOV-1998;	98JP-00325657.	
XX			
PR	16-NOV-1998;	98JP-00325657.	
XX			
PA	(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.		
XX			
XX	WPI; 2000-567637/53.		
DR	N-PSDB; AAA62000.		
DR			

XX			
PT	A protein Synanon.		
XX			
PS	Claim 1; Page 5-10; 13pp; Japanese.		
XX			
CC	This sequence represents rat p3103 protein, also referred to as Synanon		
CC	in the title of the specification. The p3103 protein contains both an SH3		
CC	domain and a PDZ domain, and was identified as being able to interact		
CC	with the protein SAPA1 via the yeast two-hybrid system. p3103 protein is		
CC	therefore thought to be involved in neuronal function, and may be useful		
CC	for the study of the human nervous system, and for the diagnosis,		
CC	prevention and treatment of various neurological disorders		
XX			
Qy	593	BEAI--TPQSAEL--ASMEI--SSRWTE--EEMETAKKGLLEHGRNWSAIARMVGSKT	646
Db	754	DEAVHKASQAKRLPPPAISLRKSMTESELEWSPWKKIEYEQQAAVPSMEKKRTV	813
Qy	647	SQCKNFYFNYKQRQNLDEILOQHLKMEKERNARRKKKAPAAASEEAAFPVVEDEME	706
Db	814	YQ-----MALNKLDEIL-----AAAQQTIS-----ASE	836
Qy	707	ASGVSGNEEMVEEAEALHASGNEVPRGECGSPATVNNSSDTESIPSPHTEAAKTQNG	766
Db	837	SPGPG-----LASLGKHPKGF-----TESSFDPHRS-----QPS	870
Qy	767	PKPPATLGADGPPGPPPPRRTS-----RAPIEPTA---SEATGAPTPPPAPPSA	817
Db	871	YDRPSEL-----PPQGLMLRQKSGIAEDDRPYLAPPAMKFSRSLVSGSDIIPPTT	925
Qy	818	PPVVPVKEEETAAAPVEEGEQKPPAAEELAVDTCKAEPEVKSECTEAEEGPAKG	877
Db	926	SPP-----EPPY-----STPPA-----	937
Qy	878	KDAEAAEATAEGALKAEKKEGSGRATTA-----KSSGAPQDSDSATCSADEVDAE	930
Db	938	-----PSSGRLTSPRGPPNPSGGPLPASSPSPDGPSPDTR	978
Qy	931	GGDKNRLSPRPSLLTPTGDPANASQKPLDLKQKQRAAIPPIQVTKVHEPP---RE	987
Db	979	GGREKSL-----YHSAALPPAH---HHPPHHHH	1005
Qy	988	DAAPTGP-----APPAPPPQNIQPSDAPQQGSSPRGKSRPAPPA	1030
Db	1006	HAPPQPHHHAHPHPPEMETGGSPDDPPRLALGPQPSLRGWRGGGSPSTGAPSPSH	1065
Qy	1031	DKEAPAAEAQKLPDPPCWTSGLPFPVPRVIVKASHAPDPSAFSYAPP-----GH	1082
Db	1066	HSSGGSG---PTQAPALRY---FQLPPR-AASAAMYVPARSGRKGLPVKQTKVGE	1118
Qy	1083	PLPLGLHDTARVLPRPPTISNPPPLISSAKH-----PSVLER-----QIGALISQGSVQL	1133
Db	1119	FQKGSIPASSTTTALPR-SEPPPAGESEKNSIPIITIIKAPTSSSGRSSQSSTEA	1177
Qy	1134	HVPYSEH-----AKAPVGVMTGLPLPMDPKKLAPFSGVQEQLSRPGQAP	1180
Db	1178	EPPTQPDGAGGGGSSPAPATSPVPPSPSPVPTPASPSGPATLDTQFGAALVGAARK	1237
Qy	1181	PESLGVPTAQEASVLRGTALGSPGSGITKGIPTRFVPSDSAITVRSITHTPADVLYK	1240
Db	1238	EGGWQNEARRSTLFLSTDAGDEGDSGLG-FGG--PFGPLRHSKSIDEG-----	1286
Qy	1241	GTITRIIGEDSPSLDRGREDSLPKGHVIEYEGKHVLSYEGGMSVTQCSKEDGRSSSGP	1300
Db	1287	-----MFSAEVYLRLESG-----GSSGGYGAYAAG-----SRYTGGSGSD	1322
Qy	1301	PHETAAPKRTYDMMGRVGRRAISSASIEGLM-----GRAIPPER-----	1339

Query Match

Best Local Similarity

Matches 384; Conservative 140; Mismatches 627; Indels 694; Gaps 84;

Sequence 2091 AA;

Score 429; DB 3; Length 2091;

Pred. No. 6.9e-15;

Db	1323	AFTSFLPPR--PLVHPLTGKALDPASPLGLAALAAERALKESSEGCTQPPPPRPPSPRY	1380
Qy	1340	-----HSFH--HLKEQHHR-----GSITQIPRSYVEAQED	1369
Db	1381	DAPPTLHHSPHSPHARHEPVLRLWGDPPARRELGYRAGLGQEKALATASPPAARS	1440
Qy	1370	YLR-----EAKLLKRGTPPPPPSRLDTEYKQALGLPLKPAHEGLVATVKEAG	1422
Db	1441	LLHRLPPTAPGVGPLLQLG-PEPTPHPGVSKAMRTAA-----	1478
Qy	1423	RSIHETPRELHTRP-LPLAPRLKEGSIOTGTLKYDTGASTGSKKHVRSLSGSPG	1481
Db	1479	-----PEE-----PERLPHURFLEN-----COARP-----PPAGTGSSTED-----GPG	1514
Qy	1482	RTFPVPHPLDVMDARALERACRYEESLKSRRPTASSGGSIARGAPVIVPELGLKQPSPL	1541
Db	1515	-----VPPSPRRVL-----PTSPTSRCNEENGLPLV-----LPPAPS	1550
Qy	1542	TYEDGAPFAGHLPRGSPVTMRPTRLQEGSLSSKASQDRKLTSTPREIAKSPHSTVP	1601
Db	1551	VDVDDG-----EFLFAEPLPPLEFNSFEK-----PESPLT	1582
Qy	1602	EHHPHISPYEHLRLGVSVDLYRSHIPLAFDPTSIPIGILDAARAYVLPRLHAPNPT	1661
Db	1583	PGPPHLP-----DPPS-----PATPLPAA-----	1602
Qy	1662	PHLYPPYLIRGYDPTAALENROTIINDYITISQOMHNTATAMAQADMLRGLSPRESSLA	1721
Db	1603	-----PPPAVAAPPT--LDSTASLTSY-----DSEVA	1629
Qy	1722	LYNAGPRGIIDLSQVPHPLVPLVPPPTGPTATAMDRLAYLPTAPQPFSSRHSSPLSPG	1781
Db	1630	TLTQGAAPAGD-----PPAPGPPAPA-----APAPPAPQ-----G	1661
Qy	1782	PHLYKPTTSSSERDRDREREKSLTSTTVEHAPIWRPCTEASSGSSGS	1841
Db	1662	P---DPPTGDSGIEVDSRSSDHPLE---TISSASTLSL-----SAEGGNTGGVA	1709
Qy	1842	GGGGSSSRPASHAHQHSPIPRQDQALQORPSVLHNTGMKIITAVEPSKPTVL---	1898
Db	1710	GGAGVAS-----GTELLDTYVAVLDAQAGSGSTPGPPVPPQLMTPS	1752
Qy	1899	-----RSTSTSSVPRPAATPPATHCPGLGTLDGVYPTLMPEVLLPKAPRVARPERPRAD	1954
Db	1753	KLGRALGTSGNLRPG-----PSGGLRDPVTPT-----SPTVS--VTGAGT	1791
Qy	1955	TGHAFILAKPPARSGLEPASPSKSGSEPRPLVPVSGHATIARTPAKNLAPHASPPPPAP	2014
Db	1792	DGLLALSACPGPSTAGVAGP---VAVEPEVPVPLPA---ASSLPKLLPWEEGPGPPPP	1846
Qy	2015	--PASASDPHREKTOG-KPFSIOELRLSLGYHSSYSPEGVPEVPVSPSPSLT--HDKG	2069
Db	1847	PLPGPLSQOASNALATVKASIISELKKLQFGGSS-TAGGALPWARGSGGSTDHSHGG	1905
Qy	2070	---LPKHLBELDKSHLEBLRKPQGVKLGGEAAHLPHLRPLPESQPSSS-----	2117
Db	1906	ASVIPERTSSLOKORLSED-----SQTSLL-----SKPSSSIFQNWPKPP	1945
Qy	2118	-PLQTAPGVKGHORVVTVAQHSIVITQDYTRHHPQQLSAPLPALYSFPFGASCPLDL	2176
Db	1946	LPLPTGSGVSSSTAAAGATSPSASSASASTR-HLQGVFEFEMRPL-----L	1992
Qy	2177	RPPPSDLYLPPDPHGAAPARGSPHSEGGKSPBNKTSVLGGGEDGIEPVSPPEGMTEPGH	2236
Db	1993	RRAPPSLLPASDH-----KVSPAPRPSL-----PILP-----	2021
Qy	2237	SRSAVYPLYRDEQTEP-----SRMGSKSPGNTSQPPAFPSKITE	2277
Db	2022	-SGPIYPLGLF-DIRSLQEGREARLTPLPSLCHHTPGYLVLEE	2064

each encoding a different MSF protein. Exons V and VI are thought to be related to the activity of the factor and are also implicated in the stability, folding and processing of the molecule. These exons are also thought to play a role in the observed synergy of MSF with other cytokines. Exons V - XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor, ie one or more of these exons may contain sequences which direct proteolytic cleavage, adhesion, organisation of the cellular matrix or extracellular matrix processing. Both naturally occurring and non-naturally occurring MSF's may be characterised by various combinations of alternatively spliced exons from this sequence, with the exons spliced together in differing orders to form different members of the MSF family. (Updated on 25-MAR-2003 to correct PN field.)

CC	XX	Sequence 1404 AA;
CC	XX	Query Match 3.2%; Score 428; DB 2; Length 1404;
CC	XX	Best Local Similarity 20.8%; Pred. No. 4.6e-15;
CC	XX	Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;
QY	DB	476 NENYKSLVRRSYRRKGSQ-----QQQQQQQQQQQQQQQMPRSPRSEKDEKE 524
DB	DB	126 SQTIKSTTKSRPKPPNKKTKKVIIEEITEHSVSENOESSSSSSSSSTIMWIKSS 185
QY	DB	525 KKEAEKE-EEPEVENDKEDLLKEK-----TDDT-SGEONDEKAEVASKRKTANSQ 575
DB	DB	186 KNSAANRELQKLLKVDKNKNRTKKTPKPPVVDGAGSLONGDPKVTPTDSTTTQHNK 245
QY	DB	576 GRRKGRTSMANEANSEAITP-QSAAELASMELESSESWTEEMETAKGLLEHGRNW 634
DB	DB	246 VTSPPKIT--TAKPINRPSLPPNSOTSKETSLTVNKETTVETKTTTNNKOTSTDGKEK 303
QY	DB	635 SAIAIRMGSKTVSQCNFYFNKYKQNLDEIILOQHLKME---KERNARRKKKAPAAAS 691
DB	DB	304 TTSAKETQSIKTSADL-----APTSKVLAKPTPKAETTTKGALITPKETPTTPK 356
QY	DB	692 EBAAPPPVVEDEMEASGVSGNEEMVEBAELHASGNEVPREGCGSPATVNNSSDTSI 751
DB	DB	357 EPASTTP-----KEPT-----KEPT-----PTTIKSAPTPK 380
QY	DB	752 RSPHTEAAKDTGONGPKPATILGADPPPGPPPTPRRTSRAPLEPTPASEATCAPTP-- 809
DB	DB	381 PAPTITKSAPTTPKEPAPTT-----KEPAPTTKEPAPTTTKEPAPTT-TKSAPTTPKE 434
QY	DB	810 PAPPSPAPPPVVPVVEKEEETAAAPVBEGBEQKPPAAAEELAVDTGKAEPEPVKSECTEE 869
DB	DB	435 PAPTTPPKPAPTTPKE--PAPTTPKEPTTPKEPAPTTKEPAPTTKEPAPTT----- 485
QY	DB	870 ABEGPAKGDAEAAEATAGALKAEKKEGSGRAITAKSSGAPQDSDSATCSADEVDEA 929
DB	DB	486 APKKPAPTTPKEPAPTTKEPAPTTKE-----PSPTTPKEPAPTTTKSAPTTPKEP-- 537
QY	DB	930 EGGDKNLLSPRSLTPTGDPANASPOKPLDLKOLKORAAAIPIQVTKVHEP-----P 985
DB	DB	538 -----APTITKSAPTTPKEP-----SPTITKEPAPTTTP 565
QY	DB	986 REDA--APTTPAPPPPPQNLOPESDAPQPGSSPRGSRSPAPPADKEAFAAEAQKLP 1043
DB	DB	566 KEAPPTPKPAPTTPKEPAPTTKEPAPTTTTPKPAAPKAPAPTTKETAPTTPKLT 625
QY	DB	1044 GPPP-----CWTGSLPFPVPPREVIVKASHPADPSAFSYPAGPHPLPLGLHDRTARVLP 1099
DB	DB	626 PTTPEKLAPTTPKEPAPTTPEELAPTTPEPTPTT-----PEEPAPT-TPKAAAPNTKE 679
QY	DB	1100 PTISNP-PPLISSAKHPSVLEROIGAISQMSVQLHVPVSEHAKP-----VGPVTMGLP 1153
DB	DB	680 PAPTTPKEPAPTTPKEPA-----PTTPKETAPTTPKGTAPTTLKEP 720
QY	DB	1154 LPMDPKPLAPFSVKOEQLSPRQAGPPBSLG---VPTAQEASVLGTLGSLVPGSGITK 1210
DB	DB	721 APTTPKKAP-----KELAPTTPKETSTSDKAPAPTT-----KGTA----- 758
QY	DB	1211 GIPSTRVPSDAITYRGSITHGTPADVLKGTITRIIGEDSPSRDLRGREDSLPKGHVITY 1270

DB	QY	759 --PTT--PKBPAPT-----TPKEPAPTTTKGTAPTTLKEPAPTTPKKPAKELAP----- 804
QY	DB	1271 EGKKGHVLISYEGGMSVTQCSKEDGRSSGPPHETA--APKRTYDMMEGRVGAISSASIE 1328
DB	QY	805 -----TTTKGPTST-----SDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETP-- 850
QY	DB	1329 GLMGRAIPP-----ERHSPHHLKEQHHRISITOGIPRSYVEAOEDYLRRBAKLLKREGTP- 1384
DB	QY	851 -----PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAPTTPK 887
QY	DB	1385 -----PPPPPSRDLTEAYKTOALGPLKLKPAHGLVATVKEAGRSIHEIPREELRHTPE-L 1439
DB	QY	888 ALENSPKPGVPTTKT-----PAATKE-----MTTAKD-----KTTERDLRTTPTT 931
QY	DB	1440 PLAPRLKEGSIT-----QGTPLKYDTCAS-----TTGSKK 1470
DB	QY	932 TAAPKMTKETAITTEKTTESKITAITTQVTSITTTQDTTFFKITTLTKTLAPKVTTTKT 991
QY	DB	1471 HDVRLISGSPGTFFPVHPLDVNADARALERCYBESLSKSRPGTASSSGS-IARGAPVI 1529
DB	QY	992 ITTTEIMNKPEET-----AKPKDRATNSKATTPKQKPKAPKPTSTTKPKT 1039
QY	DB	1530 VPGLCKPROSLTYEDHGAPFAGHLPRGSPVTMRE--PTPRLQEGSLSSS---KASQDRK 1584
DB	QY	1040 MPVRVRKPKTTP-----TPRKWTSTMPLELNPSTRIAEAMLQTTTTRPNQTPNSK 1086
QY	DB	1585 LT-----STPREIAKSPHSTVPEHHPH-----PISPYEHLRLGVSGVDLYRS 1626
DB	QY	1087 LVEVNPKSEDAGAGETPHMLLR-PHVMPPEVTPDMDLPRVFNQGIIN----- 1136
QY	DB	1627 HIPLAFDPTSPRGIPLD 1644
DB	QY	1137 --PMLSDETINICNGKPPVD 1152
RESULT 65		
AAB29773	ID	AAB29773 standard; protein; 1404 AA.
XX	AC	AAB29773;
XX	DT	28-FEB-2001 (first entry)
XX	DE	Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
XX	KW	Human MSF; megakaryocyte stimulating factor; tribonectin; alternative splicing; joint boundary lubricant; O-linked oligosaccharide; osteoarthritis; tribosupplementation; tissue adhesion inhibition; friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
OS	OS	Homo sapiens.
PN	PN	WO200064930-A2.
XX	PD	02-NOV-2000.
XX	PF	24-APR-2000; 2000WO-USO10953.
XX	PR	23-APR-1999; 99US-00298970.
XX	PA	(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX	PI	Jay GD;
XX	XX	WPI; 2001-024673/03.
DR	DR	N-PSDB; AAC81498.
PT	PT	Novel tribonectin polypeptide useful as lubricant for treating osteoarthritis, comprises O-linked lubricating moiety.
XX	PS	Claim 3; Page 7; 47pp; English.
XX	XX	

PR	18-APR-2000; 2000US-00552929.	QY	986	REDA--APTKPAPPAPPQNLQPHSDAPQOQSGSPGRKSRSPADKAEFAAQAQKLP	1043
PR	26-JAN-2001; 2001US-00770160.	Db	577	KEPAPTTTKKPAPTTKKGPAPTTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA	636
PA	(HYSE-) HYSEQ INC.	QY	1044	GDPP---CWTSGLPFPVPPREVIVKASPHADPSPAFSVPAGHPLPLGLHDTARVLP	1099
PI	Tang YT, Liu C, Drmanac RT;	Db	637	PTTPEKLPATTPKPAPTTPELAPTTPPEPTPTT-----PEEPAPT-TPKAAANPTPK	690
PR	WPI; 2001-611725/70.	QY	1100	PTISNP-PPLISSAKHPSVLERQIGAISQGMVQLHVPYSEHAKP-----VGPVTMGLP	1153
PR	Nucleic acids encoding a range of human polypeptides, useful in genetic	Db	691	PAPTTPKEPAPTTPKEPA-----PTTPKETAPTTPKGTAPTTLKEP	731
PT	vaccination, testing and therapy.	QY	1154	LPWDEKLPAPSGVKQEOQLSPRGQAGPSPESLG---VPTAQEASVLRGALSGVPGSITK	1210
PS	Claim 20; Page 573; 765pp; English.	Db	732	APTTKKKAP-----KELAPTTTKPTSTTSKDPAPTTP-----KGTA-----	769
CC	The invention relates to novel human secreted polypeptides. The	QY	1211	GIPSTRVPSDSAITVRSITHTGTPADVLVYKGTITRIIGEDSPSRDLDRGREDSLPKGHVY	1270
CC	polypeptides and antibodies to the polypeptides are useful for	Db	770	PTT--PKEPAPT-----TPKEPAPTTPKGTAPTTLKEPAPTTPKPAKELAP-----	815
CC	determining the presence of or predisposition to a disease associated	QY	1271	EGKKGHVLSYEGGMVTCQSKEDGRSSSGPPHETA--APKTYDMMEGRVGRASIASIE	1328
CC	with altered levels of polypeptide. The polypeptides are also useful for	Db	816	-----TTTKGPTSTT-----SDKPAPTTPKETAPTTPKPAPTTPKPAPTTPETP--	861
CC	identifying agents (agonists and antagonists) that bind to them. Cells	QY	1329	GLMGRAIPP---ERHSPHLKEQHHRGSIHQGIPRSYVEAQEDYLREAKLLKEGTP-	1384
CC	expressing the proteins are useful for identifying a therapeutic agent	Db	862	-----PPTTSEVSTPTTKEPTTIHKSPDESTPE-----LSAEPKPK	898
CC	for use in treatment of a pathology related to aberrant expression or	QY	1385	-----PPPPPSRDLTAYKTOALGPLKLPKPAHEGLVATVKEAGRSIHEIPRELRHTEP-L	1439
CC	physiological interactions of the polypeptide. Vectors comprising the	Db	899	ALENSPKPBGVPTTKT-----PAATKPE---MTTAKD-----KITERDLRTTPEIT	942
CC	nucleic acids encoding the polypeptides and cells genetically engineered	QY	1440	PLAPRLKEGSIT-----QGTPKLYDTGAS-----TTGSKK	1470
CC	to express them are also useful for producing the proteins. The proteins	Db	943	TAAPKMTKETATTTKTTESKITATTTQVSTTTQDTTFFKITLTKTTTLAPKVTTKT	1002
CC	are useful in genetic vaccination, testing and therapy, and can be used	QY	1471	HDVRSLLIGSPGRTPPVHPDLVMDARALACVSESLKSRPGTASSSGGS--IARGAPVI	1529
CC	as nutritional supplements. They may be used to increase stem cell	Db	1003	ITTTIMNKPET-----AKPKDRATNSKATTPKPKTAPKPKTTPKPKT	1050
CC	proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon	QY	1530	VPELGKPSQSLTYEDHAGFAGHLPGRSPVTMRB--ETPRLQEGSLSSS---KASQDK	1584
CC	and/or nerve tissue growth or regeneration; immune suppression and/or	Db	1051	MPRVKPKTTP-----TPRKWTSTMPELNPTSRIAEAMLQTTTRPNQTPNSK	1097
CC	stimulation; as anti-inflammatory agents; and in treatment of leukaemias.	QY	1585	LT-----STPREIAKSPHSTVPEHHPH-----PISPYEHLRLGVSGVDLYRS	1626
CC	AAU29510-AAU33304 represent the amino acid sequences of novel human	Db	1098	LVEVNPKSEDAGAGETPHMLLR--PHVFMPEVTPDMVDYLPVFNQGIIN-----	1147
CC	secreted proteins of the invention	QY	1627	HIPLAPDPTSIPIRGIPLD	1644
XX	Sequence 1415 AA;	Db	1148	--PMLSDEFNICNGKFPD	1163
Query Match					RESULT 68
Best Local Similarity 20.8%; Score 428; DB 4; Length 1415;					AAW78710
Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;					ID
QY	476	NYENKSLVRSYRRRKQSQ-----QQQQQQQQQQQQQQQPMRPSQEEKDEKE	524	AAW78710 standard; protein; 2063 AA.	XX
Db	137	SQTIKSTTKRSPKPNKTKTKVIESEITEHSVSNQESSSSSSSSSTIWKIKS	196	AAW78710;	XX
QY	525	KEKEAEKE-BEKPEVENDKEDLLKE-----TDDT-SGEDNDEKAEVASKRGTANSQ	575	06-NOV-2001 (first entry)	XX
Db	197	KNSAANRELQKLVKDNKNRKTTPKPPVWDEAGSLDNGDFKVTTPDSTTQHNK	256	Human protein SEQ ID NO 1372.	XX
QY	576	GRKGRITRSMANESEAITP-QQSABLASMELNESSRWTEEMETAKGLLEHGRNW	634	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	XX
Db	257	VSTSPKIT--TAKPINRPSLPNSDTSKETSITVNETTTTNNKQTSSTDGKEK	314	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	XX
QY	635	SAIARMVGSKTYSQCKNFYFNKQRQNLDEILQHKLMK-----KERNARRKKKKAPAAAS	691	tissue growth factor; immunomodulatory; cancer; leukaemia;	XX
Db	315	TTSAKETQIEKTSADL-----APTSKVLAKPTPKAETTKGPAITTPKEPTTPK	367	nervous system disorder; arthritis; inflammation.	XX
QY	692	EEAAPFPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGBCSGPATVNNSSDTSI	751	Homo sapiens.	XX
Db	368	EPASITP-----KEPT-----PTTIKSAPTPKE	391	WO200157190-A2.	XX
QY	752	PSPHTEAAXDTQONGPKPATIGADGPPGPTPPRRTSRAPITPTPASEATGATPP--	809		
Db	392	PAPTTPKSAPITPKPAPTT-----KEPAPTTPKEPAPTTPKPAPTT-TKSAPITPK	445		
QY	810	PAPPSAPSAPPVVPVVEKEEETAAPPVVEGEQKPPAAELAVDTGKAEPEVKSECTEE	869		
Db	446	PAPTTPKPAPTPKB--PAPTTPKEPTTTPKEPAPTTPKPAPTTPKEPAPT-----	496		
QY	870	ABEGPAKGDAAEAATAGALKAEKKGSGRATTAKSSGAPQSDSATSADVEDEA	929		
Db	497	APKGAPTTPKEPAPTTPKEPAPTTPKE-----PSPTTPKEPAPTTPKSAPITTPKE	548		
QY	930	EGGDKNRLLSPPSLLTPTGDPFRANASPOKPLDLKQRAAIIPIQVTKVHEP-----	985		
Db	549	-----APTTPKSAPITTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP	576		

[illegible]

1279	ILKAIQAPSNLTWNPNSFATPOTHKLDLSVVVNSGK-QSNSGATKQASPSNRRSSPGSS	1333
1984	-LVPPVSGHATIAARTPAKNLA-----PHH--ASPDPPA-PPASAGDPHREKT	2026
1338	RKTTSPGRQN-SKAPKLTLASQTNAAALLQNVELPRNVLSPTPLANPPVPGSFPNNS--	1394
2027	QSKPFSIQELESRLSGVHGSSYSPGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEGEL	2086
1395	-----GLNPQNSTVSAVAGGVVEDNKESLNPQSDCQNSQSRKEQVNIEL	1441
2087	RPQCPGVKLGGEAAHLPHLRPLPESQPSSSPLLOTAPGVKGHORVVTLAQHISEVITQD	2146
1442	KAVPAQEVKM-----VVPEDQSKKGQPSDPNKL-----PSVEENKLVSPAM-----	1484
2147	YTRHHPOOLSAPLPAFLYSPFGACPVLDLRRPP-SDLYLPPPDHGCAPARGSPHSEGGK	2205
1485	--REAPTSLSQLL-----DNSGAPNVTIKPPGLTDLEVTTP-----	1518
2206	SPEPNKTSVLGGEDGIEPVSPPEGTEPGHRSASVYPLLRYDGEQTEPSRMGSKSPGNT	2265
1519	-----VVSGED-----LKKASVIPTL-----QDLSSSKPEPSNS	1546
2266	SQPPAFSKLTESNAMVSKKOEINKKLNTNHRNEPEVNIISQGTETPFNMBAITGTGLM	2325
1547	LNLP-----HSNELCSSLVHPDELSEVS-----NVAPSIPPVMGRP-VSSSSIS	1589
2326	TYRSQAVQEHASTNMGLEAIRKALMGKYDQWEESSPPLSANAF---NPLNASASLPAAMP	2382
1590	T-----PLPPNQITVFTSNPITTSANTSAAIP	1617
2383	ITAADGRSDHTLSPGGGKAKVS-----GRP-----SSRKAK	2415
1618	THLQSLMSTVVTMPNAGSKVMYSEGSAQAQSNARPOFITPVFINSSIIQVMKGSQPS	1677
2416	SPAPGLA--SGDRPPSPSVV---HSEGDGN-RRTPLTNRWEDRPPSAGSTPEPY-----N	2465
1678	IPAAPTNTNSGLMPPPSVAVVGLPHIPQNIKFSSAPVP-----PNAISSPAPNIQTGR	1730
2466	PLTMRLOAGVMASFPFPPGGLPAGSGPLAGPHHAWDEBPKPL	2505
1731	PLVLSEATPVQLPSP---PCTSSPVV-PSHPVPVQVQKEL	1766
RESULT 70		
ABG17147		
ID	ABG17147 standard; protein; 2063 AA.	
XX	ABG17147;	
AC		
XX		
XX		
18-FEB-2002	(first entry)	
XX		
DE	Novel human diagnostic protein #17138.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
XX		
PN	WC200175067-A2.	
PD		
XX	11-OCT-2001.	
PF	30-MAR-2001; 2001WO-US008631.	
XX		
PR	31-MAR-2000; 2000US-00540217.	
PR	23-AUG-2000; 2000US-00649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Drmanac RT, Liu C, Tang YT;	
XX		
XX	WPI; 2001-639362/73.	
DR	N-PSDE; AAS81334.	
DR		

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX

PS Claim 20; SEQ ID NO 47506; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2063 AA;

Query Match 3.2%; Score 428; DB 4; Length 2063;

Best Local Similarity 18.9%; Pred. No. 7.7e-15;

Matches 408; Conservative 269; Mismatches 837; Indels 646; Gaps 98;

QY 591 NEEAATPOOSAEALAMELNESSRWTEEMETAKGLLEHGRNWSAIAVMGSKTVSQCK 650

DB 8 NLEDIYTSICSTMEDSEDFDGL--EDDDTKSDSILEDDSTIFVAFKGNIDDKD----- 60

QY 651 NFYENYKKQNLEIIL-----QQHKLKWEKER--NARKKKKAPAAASEEAFPPV 699

DB 61 ----FKWK-----LDAILKVNPLNLMHSEKSKVKQVEPWNVSRTFNIPREAAERLRLIAQ 113

QY 700 VEDDEASGVSGNEEBEEMVEAEALHASGNEVPRGECGPATVNNSSDSTESIPSPHTEAA 759

DB 114 SNNQQLRDLGILLSVQIE-GEQALNLAQNRSDQVRMNGPMGAGNSVRMEA-GFPWASGP 171

QY 760 KDTGQNGPK-----ATLGADGPPP--GPPTPPRTGRAPTEP-----TPASE 801

DB 172 GIIRMNPNATVMPFGGNYSSNMAFGNPPELQPRTPRPAQSDAMDPLLSGLHIQQQSH 231

QY 802 ATGAPTPPPAPPSP-----SAPPPVVPKEEKEETAAAPVVEGEQKPPAAEEELAVDTG 856

DB 232 PGSGLAPPHHPQVSVNRMKNPAPFQIQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 291

QY 857 KAEPPVKSECTEAEAGPKAGDAEAAEATAGALKAEKKGSGRATTA---KSSGAPQ 913

DB 292 QQHQQQQQPGIRPQFTAPTQVPPVPGWNLPSGALQPPPAQGSLSGTMTANQGWKKAPLP- 350

QY 914 DSDSSATCSADEVAEGDKNRLISPRPSLLTPTGDRNANASPOKPLDLKOLKORAAAI 973

DB 351 -----GFMQQQLQARPSLAT----- 365

QY 974 PPIQVTKVHPREDAAATKPAPPPPPQNLQPSDAPQ--PGSSPRGSKSRPAPADK 1032

DB 366 --VQT-----PSHPPPPYPFGQQQASQANTNFPQMSNPGQFTAP----- 402

QY 1033 EAFAAEAOQLPGDPPCWTSGLPFPVPPREVIVKASPHADPPSAFYS-APPGHPLPLGLHDT 1091

DB 403 -----QMSLQGGP-----SRVPTLQOQPHLTNKSPP-ASSPSSFQOGSPASSPT---VNQT 449

QY 1092 ARPVLPRPTTINP-----PPLISSAKHPSVLERIQIGAISQMSVOLHVPVSEHAKAPV 1145

DB 450 QQOMGPRPPQ--NNPLPQGFQQQVSSPGRNPMV---QQGNVPPNFMVMOQQPPNQ----- 499

QY 1146 GPVTMGLPLPMDPKKLAP--FSGVKORQLSPRQAGPPPSLG-VPTAQEASVLRGTALGVS 1203

DB 500 GPQSLHPLGGLGMPKRLPPGFSA-----QGANPNFMQGOVP-----STTATT 540

QY 1204 PGGSITKGPSPTRVPSDSAITYRGSIITHTPADVLYKGTITRIIGEDSPSLDRDREDSL 1263

DB 541 PGNIS---GAP--OLQANQNVQHAGGAGPPQNM-----QVSHGPPNMM 580

QY 1264 PKGHVIEYEGKGHVLSYEGGMSVTQCKEDGRSSGPPHETAAPKRTYDMMEGRVGRAIS 1323

DB 581 QPSLMGIHGNMNMNQAGTSGVPQVNLNMQGPPQGGPP----- 618

QY 1324 SASISGLMRAITPRRHPHHLKEQHHRGISTQIGIPRSYVEAQEDYLREAKLL--KR 1380

DB 619 -SQLMGMHQIIVP-----SQGQMVQO---QGTLPQNPMILSRAQ---LMPQGMVNPSP 667

QY 1381 EGTTPPPP---PPSRDLTE-----AYKTOALGP---LKLKPAHEGLVATVKEAGRSI 1425

DB 668 QNLGSPQBMTPPKQMLSQGQPMWAPHNMGMPOGOVLLQONPMIEQIMTNMQGNKQO 727

QY 1426 HEIPREE--LRHTPELPLAPRLKEGSITQGTPLKYDTGASTTGSKGDHVRSLIGSPGT 1483

DB 728 FNTQNSQNVMPGPAQIMRGPTNMQNMVQFTGQMSQMLPQOGPVNNSPSQVMGIQGOV 787

QY 1484 FPPVHPLVDADARALERACYEESLKSRRGTASSGGSIARGAPVI-----VPELKP 1536

DB 788 LRPPGSPHMAQOH-----GDPATTANNVSLSQMMPDVSIOQTNNVPHVQA 835

QY 1537 RQ-----SPLTYEDHG---APPAG-----HLPRGSPVTMREP--TPRLQEGSL 1574

DB 836 MGNASASGNHPSGKMSFNAPFSGAPNGNQSCQNGPVPNVKDVLTLSPLLVNLLQSDI 895

QY 1575 SSKKASQDRKJTS-----PREIAKSPHS-TVPEHHPHPSIPYEH--LLRGVSGVD 1622

DB 896 SAGHFGVNNKQNTNANKPKKKKPKKKNSQODLNTPTDTPAGLEADQDPLPQEGQIS 955

QY 1623 LYRSHIPLAFDPTSTPRGI---PLDAAAAYLP---RHLAPNPTYPLYP--PYLIR--- 1671

DB 956 LDNSG-PKLPEFSNPPGYPSPQVPEQRFLQOQMPQOLMQHVAPPPOPPQOQPOQPOQOQ 1014

QY 1672 ----GYPTAALENRQTIINDYITTSQQMHNHTATAMAQADMLRG--LSPRESSLALNAA 1726

DB 1015 PPPSPQSQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1074

QY 1727 GPRGIIDLSQVPHLPVLPPTP-----GTPATAMDRLAYLPTAQO--PPSS----- 1770

DB 1075 SVPMVSLQG---PASVPPSPDKQRMPPMVPNTPLGNSNRKMWYQESQNPSPSSPLAEMA 1130

QY 1771 ----RHSSSPSLPCGGPHTLTKPTTSSSERDRDRDREREKSIILTSTTTVEHAP 1825

DB 1131 SLPEASGSEAFSVPGGPNMFSHV-----VLPQNQLMTGP 1166

QY 1826 IWRPQTEOSSSGSSGGSSSSRSPASHSHAHQSP--ISPRTOALQORPSVLHNTGMK 1884

DB 1167 --KPGSPSLATQATPOQPPVNSLPSHGH--HFPNVAAPTQTS---RPKTPNRSAPR 1218

QY 1885 GIITAVEBSKTVLRSTST--SPVRPAATFPPTHCPGLGTLGDVYPTIMEPVLLPKAP 1943

DB 1219 PYPQTPNNRPPSTPEPSEISLSPER-----LNASIAGLFP-----P 1254

QY 1944 RVARPERADTGHAFKAPARSGLEPASPSKSGSEPRPLVPPVSGHATIAITPAKNLA 2003

DB 1255 QINPLPPRPNLNRGF-----DQOGLNPTTLKAIQAPSNLTMAPSNFAT----- 1299

QY 2004 PHASPDPPAPASADPHREKTSKPFISQIELELRSLGYHSGSSYSPGEGVPSPVSPSS 2063

DB 1300 -----PQTHKLDVVVN-----SGKQNSGATKRAGPSNS-- 1329

QY 2064 LTHDKGLPKHLEELDKSHLEGELRPQKQPVKLGGEEAHLPHLRPLPESQSSSPLLQTA 2123

Db	2222	-----EARGKPG-PQKPP-----TEAD-----KNGMKRSP-----	2247
Qy	2308	QPCTEINFMPAITGTLMTYRSQAVQH-----ASTNMGLEAIRKALMGYDQWEE\$PPL	2363
Db	2248	-----SATGQ\$F\$R\$T\$A\$L\$P\$K\$S\$L\$C\$S\$S\$F\$P\$T\$R\$A\$G\$R\$E\$A\$A\$S\$D-----	2287
Qy	2364	SANAFNPLNASASLPAAMPITADGRSDHILTS\$G\$G\$K\$-----AKV\$G\$R\$P\$S\$R-----	2412
Db	2288	-----T\$S\$A\$K\$A\$G\$G\$M\$E\$L\$P\$A\$P\$N\$R\$D\$H\$K\$A\$O\$P\$A\$G\$E\$G\$R\$T\$H\$M\$T\$K\$D\$S\$L\$P\$S\$F\$R\$V\$T\$U\$P\$L\$E\$S	2339
Qy	2413	-----KAK\$P\$A\$P\$A\$G\$L\$A\$G\$R\$D\$R\$P\$P\$S\$V\$V\$H\$S\$E\$G\$D\$C\$N\$R\$R\$T\$P\$L	2445
Db	2340	HH\$P\$D\$N\$T\$M\$G\$A\$H\$R\$D\$R\$A\$L\$S\$V\$T\$A\$T\$V\$G\$T\$K\$D\$P\$A\$-----Q\$P\$P\$A\$R\$Q\$N\$V\$G\$R\$D\$V\$T\$K\$P\$S\$P\$A	2394
Qy	2446	TNRVWEDRPSS 2456	
Db	2395	PN---TDRPIS 2402	

RESULT 74
AAU03503

ID AAU03503 standard; protein; 2523 AA.

AA
AC
AAU03503;

DT 12-SEP-2001 (first entry)

Human protein kinase #3.

Human; protein kinase; PTK; STK; cancer; cardiovascular disease;
metabolic disorder; immune related disease; neurological disorder;
neurodegenerative disorder; inflammatory disorder; infectious disease;
reproductive disorder.

OS Homo sapiens.

AA
PN
WO200138503-A2.

31-MAY-2001.

22-NOV-2000; 2000WO-US032085.

AA
PR 24-NOV-1999; 99US-0167482P.

AA (SUGE-) SUGEN INC.

PI Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R:

FI Flanagan P, Clary D;
XX

DR WPI; 2001-343950/36.
DR N-PSDB: AAS06703.

XX
PT
Nucleic acids and

PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and neuronal-associated diseases, and microbial infections.

PS Claim 7; Fig 2; 433pp; English.

AAU03501-AAU03557 represent novel human protein kinases #1-57. The novel protein kinases have been identified as members of the tyrosine or serine/threonine kinase (PTK and STK) families. The polynucleotides encoding protein kinases and the polypeptides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate kinase expression. For example, they may be used to treat cancers (especially cancers of haematopoietic origin), cardiovascular disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes), immune related diseases (e.g. rheumatoid arthritis), neurological disorders (e.g. schizophrenia), neurodegenerative disorders (e.g. Parkinson's disease), inflammatory disorders (e.g. asthma), infectious disease (e.g. HIV) and reproductive disorders (e.g. infertility). Additionally, polynucleotides encoding protein kinases may be used for gene therapy and as DNA probes in diagnostic assays. The protein kinase polypeptides may be used as antigens in the production of antibodies

CC against the protein kinases and in assays to identify modulators of
CC protein kinase expression and activity
XX
SQ Sequence 2523 AA;

Query Match 3.2%; Score 424.5; DB 4; Length 2523;
Best Local Similarity 19.8%; Pred. No. 1.6e-14;

38	GLLEYOHH-----SRDYASHLSPGSI-----IQORRRPSSLSEFPQGNERSOELHURPE	87
309	GVLSFTHHQIIEIARCDLOKSHOGLTYSRYFLELOHKDKLL-----QEAHDRSE	358
88	SHSYLPGLKSEMEFTESKPRELELPDPLLRSPPLATQOPAGSEDLTKDRSLTGKLEP	147
359	S-----GELAFIKOLVRKILIV-----IARPARLEEC-----LE-	387
148	VSPSPPHPTDPE-----LELVPPRLSKEELIQNMDRVDRITMVEQOISLKKKQOOLEE	202
388	-----PDPEFFYLLAEAAHAKGEGQIKT--DIPRYI-----ISQLGLNKOPLEE	431
203	EA-----AKPPPEKPVSPPPPIESKHSRSLVQIYIDENRKAKAAHRILBGLGPQVE	253
432	MAHLGNYDSGTAEPTPETDESVSNNASLKL-----RKPRE-----	467
254	LPLYNQPSDTQVHENIKI--NOAMRKKLLYFKRRNHAKQWKQKFCQRY-----DOLM	306
468	-----SD-----FETIKUISNAYG---AVYFVRHKESQRFAMKINKONILRNQIQ	513
307	EA--LEKKVERIENPPRRRAKESVREYYEKQFPEIRKQRELOERMOSRVG-----QRGS	359
514	QAFVERDILTFAPNP-----FVVSWMYCSFETRHLCHMVEYEGDCATLKMNM	562
360	G--LMSGAARSEHEVSEIIDLSEQMLEKQMJOLAVIPMLYDADQORIKFINNM-----G	414
563	GPLPVDWMARM--YFAETVLALBYLHNYGIVHRLDK--PDNLLVTSMGHIKLTFGLSKVG	618
415	LMADPMKVYK-----DROVNMWSEQEKETFREKFMQHPKNGFGLIASFLERKTV	463
619	LSMVTNLVYEGHIEKDAREFLDKQVCGTPEYIAPENVILRQGYGKPDVMAWGIILVEFLV	678
464	ABCVLYYLIT-----KNNE-----NYKSLV-----RRSRRRGKSQOQQQ	498
679	G--CVPPFGDTPEELFGQVISDEINWPEKDEAPPDQADLITLLLRQNPLERLTGGAYEV	737
499	QOQOQOQOQOQOQPMPPRSOEEKDEKEKEKAEKEBEKPEVENDKEDLLKKTDDTSGED-	557
738	KQHRFRSLDWNLSLRQKAEFIQLESEDDTSYFDRSEKYHHME---TEEEDDTNDEDF	794
558	NDEKEAVAS--KGRKTANSQGRKGRITRSMANEANSEEAITPQOASAELASME--LNESS	613
795	WNEIQFSCSHRFKVSFSS-----IDRITONGABE--KEDSDVKTSITLPTSTETLSWSS	848
614	RWTE--EEMETAKGLLEHRNWSAIARMVGSKTVSQCKNFYFNYKKRQNLDELLOQHKLK	672
849	EYXEWQOLSTSNSSDRESNR-----HKLSSGL--LPKLAIIS	882
673	MEKERNARKKKKAPAAASEEAAFPVVEDEWEASGVSGNBEEMVEEAEALHASGNEV-	731
883	TEGEOD---EAAASCPODPHEEPCKPALPPEE-----CAQEPEVTTTASTISSTLISVG	933
732	-----PRGECGPATVNNSSDTEIIPSPH-----TEAAKDTGONGPKPPATL	773
934	SPSEHLDDQINGRSECDV--STONSSKPSSEPAHMARQLESTEKKKISG---KVTKSL	987
774	GADG-----PPPGPPTPPRRTSRAPTEPTPASEATCAPTPPPAPPSPSPAPPVP	821
988	SASALSMLIPGDMFAVSPGLSPMSPHSLSS-----SDPSSRRDSSPREDSSAASPHQPI	1042
822	VPKESKEEB--TAAAPPVEEGE-----EOKPPAAE-----	849
1043	VTHSSCKNGFTIRAIRVYVVGSDIYTVHHIIVNVBEGSPACOAGLKAGDLITHINGEPV	1102

FT Misc-difference 121 /note= "Lys encoded by GAC"
 FT Misc-difference 122 /note= "Asp encoded by AAG"
 FT Domain 448..543 /note= "SH3 domain"
 FT Misc-difference 532 /note= "Phe encoded by CCT"
 FT Domain 587..684 /note= "PDZ domain"
 XX WO200078921-A2.
 XX 28-DEC-2000.
 XX 23-JUN-2000; 2000MO-US017322.
 XX 24-JUN-1999; 99US-014071SP.
 XX (UJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 XX (MASS-) MASSACHUSETTS GEN HOSPITAL.
 XX Worley P, Tu JC, Xiao B, Sheng M;
 XX WPI; 2001-102712/11.
 XX N-PSDB; AAF24877.
 XX New Shank polypeptide for identifying Shank protein modulator compounds
 XX used to treat Alzheimer's disease and stroke, comprises an ankyrin
 XX domain, SH3 domain, PDZ domain, proline-rich domain, and SAM domain.
 XX Claim 4; Fig 2A; 118pp; English.
 XX The present sequence represents a rat Shankla polypeptide. The
 XX specification also describes Shank2 and Shank3a. The Shank polypeptides
 XX have an ankyrin domain, SH3 domain, PDZ domain, proline-rich domain, and
 XX SAM domain. Shank polypeptides play a significant role in the post
 XX synaptic density cytoskeleton. Compounds that modulate Shank proteins are
 XX useful for treating a disorder associated with glutamate receptors such
 XX as epilepsy, glutamate toxicity, disorders of memory, disorders of
 XX learning, stroke, schizophrenia, Alzheimer's disease, tissue degeneration
 XX and disorders of brain development. A disorder associated with Shank
 XX protein activity such as cardiac disorder, disorder of musculature,
 XX vasculature disorder, neurological disorder, psychiatric disorder, renal
 XX disorder, uterine disorder or a disorder of bronchial tissue is also
 XX treated by a compound that modulates Shank protein activity
 XX Sequence 2087 AA;
 Query Match 3.2%; Score 422; DB 4; Length 2087;
 Best Local Similarity 20.8%; Pred. No. 1.7e-14;
 Matches 378; Conservative 141; Mismatches 619; Indels 682; Gaps 83;
 QY 658 KRQNLDEILQHKLMKEKERNARRKKKAPAAASEEAPPPVVEDEME---ASGVSGNE 714
 DB 686 RHFDME-----AVHKASQQAQLPPALSLSKSMTSELEMEYEQQAAPVSMK 738
 QY 715 EEMVEEA-----EALHAGSNEVPRGECGPATVNNSSD-----TESIPSPHTEAAK 760
 DB 739 KRTVYQMALNKLEILAAQAQITISASESPGGLASLGKHKRPKGFATSSFDPHRS--- 796
 QY 761 DTGQNGKPKPATLGADGPPPGPPPTPRRTS-----RAPIETPA---SEATGATPPPA 811
 DB 797 ---QPSYDRPSFL-----PPGGLMLRQKSI GAIEDDRPYLAPPAMKFSRSLSPGSEDI 848
 QY 812 PPSGAPPPVPKKEKEEBETAAAPVVEEGEEQKPPAAABELAVDTCKAEPEVKSECTEEAE 871
 DB 849 PPTPTSP-----EPY-----STPPA----- 866
 QY 872 EGPAGKDAEAAATAEGALKAEKKGSGRATTA-----KSSGAPQDSSSATCSAD 924
 DB 867 -----PSSSGLRTPRGGFPNPSGGPLPASSPSSFDGP 901

QY 925 EVDEAEGDKNLLSPRSLLTPTGPRANASPOKPLDLKQLKQRAAALPIQVTKVHEP 984
 DB 902 SPPTDTGGGREGKSL-----YHSAALPPAH-----HHP 928
 QY 985 P---REDAAPTTP-----APPAPPPQNLQPEDSDAPQOQSSPRGKR 1024
 DB 929 PHHHHHAPPQPHHHHAHPHPPEMETGGSPDDPPRLALGFPQSLRGWRGGPSPSTSG 988
 QY 1025 SPAPPADKEAFAAEAQKLPDPCWTSGLFPVPPREVIVKASPHAPDPSAFYAP----- 1080
 DB 989 APSPHSSSSGSSG---PTQAPALRY---FOLPFR-AASAAVYVPASGRGKGLPVKQ 1041
 QY 1081 ----GHP-----LPLGLHDTARPVLPPTTISNPPPLISSAKH-----PSVLER-----QIG 1123
 DB 1042 TKVEGEPOKGSIPSSASSPIS-PALPK-----SEPPAGPSEKNSIPITIIIKAPSTNSG 1096
 QY 1124 AISQKMSVOLHVPYSEH-----AKAPVGVMTMGLPLMPDKPKLAPFGVKQE 1170
 DB 1097 RSSQGSSTEAEPTQPDGAGGSSSPAPATSPVPPSPSPVPTTASPSGATLDTSQF 1156
 QY 1171 QLSPRGQAGPPESLGVPTAQAEASVLRTALGSPVGSITKGIPTSTRVPSDSAITVRSIT 1230
 DB 1157 GAALVGAARREGQWNEARRSTLFLSTDAGDEGDSGLG-PGG--PPGPRLRHSKSID 1213
 QY 1231 HGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEYEGKGVLSYEGGMSVTQCS 1290
 DB 1214 EG-----MFSAEVYLRLESG-----GSSGGYGAYAAG-----S 1241
 QY 1291 KEDGRSSGPPHETAAPKRTYDMGRVGRVRAISSASIEGLM-----GRAI 1335
 DB 1242 RAYGGSGSSAFTSFLPPR--PLVHPLTKALDPASPLGLAALAAERALKESGEGTPO 1299
 QY 1336 PPER-----HSPH--HLKEQHHR-----GSIQTQGI 1359
 DB 1300 PPRPPSPRYDAPPTLHHHSPHSPHARHEPVRLWGDPPARRELGYRAGLSQEKAL 1359
 QY 1360 PRSYVEAQEDYLRR-----EAKLLKREGTPPPPPSRDLTEAYKTOALGPKLKPAHE 1412
 DB 1360 TASPPAARRSLHLRPLPTAPGVGPLLQGL-PEPTPHGVSKAWRTAA----- 1407
 QY 1413 GLVATVKEAGRSIHEIPREELRHTE-IPLAPRPLKEGSIQTGTPKLVDTGASTTGSKKH 1471
 DB 1408 -----PEE-----PERLPHVRLEN-----CQARD-----PPAGTRGSSTE 1439
 QY 1472 DVRSLLIGSPGRTPFPVHPLDMADARALACRYESLSKSRPCTASSSGSISARGAPVTP 1531
 DB 1440 D-----GPG--VPPPSPRVL-----PTSPSPRGNEENGLPLIV- 1472
 QY 1532 ELGKPROSPLTYEDHGAFFAGHLPRGSPVTWREPTRLQEGSLSSSKASQDRKLTSTPRE 1591
 DB 1473 ---LPPAPSPVDVDDG-----EFLFAEPLPPPLEFSNGFEK----- 1505
 QY 1592 TAKSPHSTVPEHHHPISPYEHLRGVGVLYRSHIPLAFDPTSI PRGILDAAAAYYL 1651
 DB 1506 ---PESPLTPGPHPLP-----DPPS--PATFLPAA----- 1531
 QY 1652 PRHLAPNTYHLVPPYILIRGYPDTAALENRQTIINDYITQSQMHNTATAMAQADMLR 1711
 DB 1532 -----PPPAVAAPPT--LDSTASSLTSY----- 1553
 QY 1712 GLSPRESSLALNYAAGRGIIIDLSQVPHLPVLVPTPTGTATAMDRLAYLTAPQPFSSR 1771
 DB 1554 ---DSEVATLTQGAAPAGD-----PPAGFPAPA---APAPPAPQF----- 1589
 QY 1772 HSSPLSPGGPHTLTKPTTSSSERDRDRERDREREKSIILTSITTTVEHAPWRPCT 1831
 DB 1590 -----GP---DPPPGTDSGIEEVDSRSSDHPLE-----TISSASTLSSL-----SA 1628
 QY 1832 EQSSGSSGSSGGGSSSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKGIIITAVE 1891
 DB 1629 EGGNGTGVAGGGAGVAN-----GTLELDTYVAYLDGQAFGGSGTGP 1671
 QY 1892 PSKFTVL-----RSTSTSPVRPAATFPFATHCPCLGGLDGVYPTLMEPVLLPKAPR 1944

Db 1672 PYPQMTSKURGRALGTSGNLRPG-----PSGLRDEVTPT-----SPT 1712
QY 1945 VARPERPRADTGHAFKAPARGLPSPKSGSEPRPLVPVPSGHATTIARTPAKNLAP 2004
Db 1713 VS--VTGAGTDGLLALSACGPGTAGVAGP---VAVEVEVPVPLPA--ASLPRKLLP 1765
QY 2005 HHASDPDPAP--PASASDPHREKTQS-KPFSIQELRLSLGVHSGSYSPGVEVPSPVS 2061
Db 1766 WEBGPGPPPPPLPGPLSQOASALATVKASIISELSSKLQOQCGSS-TAGGALPWARGGS 1824
QY 2062 PSIT--HDKG---LPHLBEELDKSHLEGLRPQPGPVKLGFAAHLPHLPLPESQPS 2116
Db 1825 GGSTDHGGGASYIPRTSLQRLSED-----SQTSL-----SKPSS 1864
QY 2117 S-----PLQTAGPVKGHORVVTLAQHISEVITQDTRHHPQQLSAPLPAPLYSF 2166
Db 1865 SIFQWPKPLPLPLPTGSGVSSSTAAPGATSPASASASTR-HLQGVFEMRPLP--- 1920
QY 2167 PGASCPVLDLRRPPSLYLPPDPHGAPARGSPHSEGGKSPENKTSV----- 2214
Db 1921 -----LRRAPSPSLPASDH-----KVPAPRPSLPLPSGPIYPGL 1958
QY 2215 -----LGGEGGIEPV-SPPE-----CMTEPGHRSANVYPLLRYDGEQTEPSR 2256
Db 1959 FDIRSSPTCGAGGSTDPFAPVFPVPPHPIGSGLGALSGASRS-----LSPTR 2006
QY 2257 MGSKSPGNT--SOPPAFFSK 2274
Db 2007 LLSLPPDKPFGAKPLGFWTK 2026

RESULT 76

ABO44406
ID ABO44406 standard; protein; 2157 AA.

XX AC ABO44406;
XX DT 26-SEP-2003 (first entry)

XX DE Novel human protein kinase #26.
XX KW Human; kinase; enzyme; cosmetic application; nutraceutical application.

XX OS Homo sapiens.
XX PN US6541252-B1.

XX PD 01-APR-2003.
XX PF 14-MAY-2001; 2001US-00854856.

XX PR 19-MAY-2000; 2000US-0206015P.
XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Walke DW, Hilbun E, Donoho G, Turner CA;
XX DR WPI; 2003-575927/54.

XX N-PSDB; ACH03789.
XX PT New nucleic acid encoding novel human proteins, useful in cosmetic and
XX PS nutraceutical applications.

XX PS Disclosure; Page; 11pp; English.
XX CC The invention relates to a new isolated nucleic acid encoding a novel

XX CC human protein kinase. The nucleic acid is useful in cosmetic and
XX CC nutraceutical applications. The present sequence represents the amino
XX CC acid sequence of a novel human protein kinase. Note: The sequence data
XX CC for this patent did not from part of the printed specification but was
XX CC obtained in electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html?DocID=6541252B1

XX SQ Sequence 2157 AA;
Query Match

Best Local Similarity 3.2%; Score 422; DB 7; Length 2157;
Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;

QY 147 PVSPPSPHTDPELELVPPRLSKEELIONMDRVREITVWEQIQISKKKQQOLEEAAK 206
Db 39 PLSLPQPSIPAAVPSQAPPEPREETV-----TATATSQVAAQPPAAAPAGQAVA 89
QY 207 PREP-----EKPSPPPIESKHSILVQIYIDENRKAAEAHRIEGLGQVDELPLY 257
Db 90 GPAPSTVPSSTKDRPVSPSL-----VGSKEBPPPA 121
QY 258 NQPSDTRQYHENIKINQAMRKLLILYKRRNHARKWKOKFCORYDQLEALMEALKKVERIE 317
Db 122 RSGG-----GSAKEPQERSQQDDI-EEETKAVGMS 155
QY 318 NNPRRAKESKV-REYEEKOFFEIRKOR-----ELQERMOSRVQQRGSLMSAARSE 369
Db 156 NDRFLKFDIEIGRSGFTVYKGLDTEVTVEVAVWCSELQDRK-----LTKSERQRF 205
QY 370 HEVSEIIDGLSQENLEKQMRQLAVIPMLYDADQORIK-----FINNGLM-ADPMKYK 424
Db 206 KEEAEMLKGL-QHPNIVR-----FYDSWESTVKGKCIVLVTELMTSGTLTKYL 253
QY 425 DR-----QVMNMWSEQEKETPREKFMHPK-----NFGLIA 455
Db 254 KRFKWKIKILRSWCQIILKGLFHLTRTPPIIHRDLKCDNIFITGTSVKIGDLGL-- 311
QY 456 SFLEKRTVAECVL-----YYLTKNENY-----KSLVRRSYRRRGKSQQOQQQQQ 501
Db 312 ATLKGASFAKSVIGTPEFWAPMEYBEKYDSDVYAFGCMLEMATSEVPYSECQAAQI 371
QY 502 QOQOQOQOQP-----MFRSSQ-----EKDEKEKEKEKEKEKEKEKEKEDEKDLK 547
Db 372 YRRVTSGVKSPASFDKVAIPEVKEIIEGCIROKNDERYSIKDLNHAFFQETGVRVELAE 431
QY 548 EKTDDTSGEDNDEKEAVASKRKTANSQGRKRITRSMANEANSE-EAITPOQSAELAS 606
Db 432 E-----DGEKIAIKLWLRIBEDIKLKGKYKDNEAIFPSFLERDVPDVAQ--- 478
QY 607 MELNESSRWTEBEMETAKKGLLEHGRNWSAIAARMVGSKTYSQCKNFYFNFKRQNDLIL 666
Db 479 -EMVESGYVCEGDHKTMAKIKDR-----VSLIK-----RKREQRQLVR 516
QY 667 QOHLKMEKERNARKKKA-----PAASEEAAPPPVVEDEMEASG 709
Db 517 EEQEKKEKQESSLKQOQVEOSSASQTGKOLPSASTGIPTASTTSASVSTQVEPEPEA-- 574
QY 710 VSGNEEMVEEAEALHASGNEVPRGCGSPATVN--NSSDTEISIPSPHTAAKDTQNGP 767
Db 575 -DOHQOLYQQPSISVLSDGTVDGCGSVFTESRVSSQCTVSYGSH-EQAHTGTVP 632
QY 768 KPPATLGADGPPPG--PPT-----PPRRTSRAPTEPT-PASEATGAPTPPPAPPS 814
Db 633 HIPSTVQAQSPHGVTVPSSVQGIQQTAPPOQTVOYLSQTSSEATTA-----QPV 687
QY 815 PSAPPPVPVKEEKEEBEETAAAPVE--EGREKPPAAEE--LAVDTGKASEPVKSECTEEA 870
Db 688 QFOAPQVLPQVAGKOLPVSPVPTIQGEPQIPVATQPSVVPVHGAFLPV----- 739
QY 871 BEGPAKGDAEAAEATAEGALKAEKGGSGRATTAKSGAPODSSATCSADEVDAE 930
Db 740 -----GQPLPTPL-----LPQYFVSQIPISIPHTVSTAQ 767
QY 931 GGDKNRLLSRPSLLTPTGDPANASFPKPLDLKOLKQFAAAIP-----PIQVTKVHEP 984
Db 768 TG-----FSSLPTWA-----AGITQPLLTASSATTAAIPGVSTVVPSPQLLQ 814
QY 985 PREDAAPTKAPAPPAPPNOLQEPESDAPQOQSSPRGKRSRPPADKAEFAAEAKLPG 1044

815 VTQ-----LFSQVHPQLQF-----AVQSMGI-----PANL-GQAAEVLSSG 851
1045 DPPCWTSGLPVPVPPREVIVKASPHADPSAFSVAEPGHPLPLGLDHTARPVL--PRPPTI 1102
852 D---VLVQGGPPRLPPOY-----PGSNIAPSSNVASCIHST---VLXPPMPTVEV 896
1103 SNPPPLISSAKHPSV---LEROIGALISQMSQVQLHVPYSEHAKAPVGPVTMGLPLPMDPK 1159
897 LATPGYFPTVQVYVESNLVPMGGV--GGQVQSQPGSLAQAPTSSSQAV-----947
1160 KLAPEFSGVQEOQLSPRQAGPESLGV-----PTAQEASVLRGALGVPGSGITKGI 1212
948 -LESTQGV-----SQVAPAEVAVAQOATPTTLASSV--DSAHSDVASG-MSDG- 994
1213 PSTRVPSDAITYRGSI THGTADVLKYGTIRIIGEDSPSLDRGREDLSKPGHVIYEG 1272
995 -NENVPSSG-----RHEGRTRKHYRKSVRSRHKTSRPKLRILNVS 1038
1273 KKGHVLSYEGGMSVTQCSKE-----DGRSSGPPHETAAPKRTYDMMEGRVG 1319
1039 NKGD-----RVVEQCQLETHNRKMVTFKFDLDGDN-----PEEIAIWNNDFI-----1081
1320 RAISSASIEGLMGRAI PPRHSPHHLKQOHHTRGSI TQIGIPRSYVEAQEDYLRRRAKLJK 1379
1082 LAIERESFVDQVREII---EKADEMLSEDSVVEPEGDGL--ESLQKDDYGFSGSKLE 1136
1380 RSGTTPPPPPSRDLTEAYKTQALGPLKAPAEGLVATVKEAGRS--IHEI PREELRHTP 1437
1137 GFQKQIPASSM-----POQIGIPTSSLQVHVHSAGRRFIVSPVESRLRESK 1184
1438 ELFLAPRPLKEGSI TQGTPLKYDTGASTTGSK-----KHDVRSL-----1476
1185 VFP-----SEIT-----DTVAASTAQSPGMNLSHASSLSLQQAFLSELRAQMT 1229
1477 -----IGSGRTPPPVHPDWMADARALERACYEBSLKSRPGTASSGSGSIARGAP- 1527
1230 GENTAPPNFSTGTPFPVVP-----FLSSIAGVPTTAAATAPVATSP 1275
1528 -----VIVPELCKPRQSL--TYEDHGAFFAGHLPRGSPVTMREPTPLRQSGSLSSKA 1579
1276 NDISTSVIOSEVTVTEEGIAGVATSTGVVTGGGL-----PIPVSESPVLSSVV 1325
1580 SQDRKLTSTPREIA---KSPHSTVPEHPHPHPI SPYEHLLRGVSGVDLYRSHIPLAFDPTS 1636
1326 SS-----ITIPAVVISITSPSLQVPTSTSEIV-----VSSITALYPS-----1362
1637 IPRGIPLDAAYLPHRIAPNPTYPHLYPPYLI-----RGYPT 1676
1363 ---VTVSATSASAGGSTATPGK-----PPAVSQQAAGSTTVGATLTSVSTTSPST 1413
1677 AALENRQ-----TIINDYITSQMHNTATAMA-----QRA 1707
1414 ASQLSLSSTSTPTLSTVVSASHLSLTKSHSSTGLAFSLASAPSSSSPCAGVSSVI 1473
1708 DMLRGLSPR-----ESSIALNYAAGPRGII DLDSQVPHLPVLVPPPTGTPATAMRLAYL 1761
1474 SQPGLHPLVPSVIASTPILQAAGTSTPLLPQVPSIPLVQPVANPAV--QQTLL 1530
1762 PTAPQPFSSRHSSPLSPGCP--THLTKPTTSSSERDRDRDREREKSLTSTTT 1820
1531 HSQQPQ-----ALLPNQPHTHCP-----EVDSDTQPKAPGIDDIKT 1566
1821 VEHAPIWRPCTBQSSSGSSGGGGSSRRPASHSHAHQHSPISTPTQDALQRPVSLHN 1880
1567 LEE-----KURLSEHSSGA-----QHASVSLETS-----LVIES 1598
1881 TGMKGI-ITAVFSPKPTVLRSTSTSSPVPAATFPFATHCPGLGTLGVYPTLMEFVLLP 1939
1599 TVTPGIPTTAVAPSK--LLTSTT-----STCLPPTNLPLG-----1631
1940 KEAPRVARERPADTGHAFLAKPPARSGLEPASPSSKSEPRPLVPVPSGHATTARTPA 1999
1632 ----TVALPVTVPVTPGQVSTPVSTTTSGVKPGTAESKPLTKAPVLVPGTLPAGTLP 1687

2000 KNLAPHASPDPPAPPASDPHREKTS--KPFSTQELRLSLGYHSGSYSPGVEPVSP 2058
1688 EQL-----PPFGPSL-----TOSQOLEDLDLAQLR-----RTLSEXTITVTS 1726
2059 VSSPSLTHDKGLPKHLEELDKSHLEGELRPQGPVKLGGEAAHPLRLPLPE---SQPS 2115
1727 V-----GPVMAAPTA-ITEAGTQPKQGVSVK 1753
2116 SSPILQTAGV---KGHQRVVTLAOHISEVITQDTRHHPOOLSAPLPAPLYSPGASC 2171
1754 EGPVLATSSGAGVFKGGRFOVSAA-----DGAQKEGKNKSEDAKSVHFESTSES 1804
2172 PVLDLRRPPSLDLYLPPP-----DHGAPARGSPHSEGGKRGSPENKTS-----2213
1805 SVLSSSSPESTLVKPEPNGIITPGISSDVPESAHKTTASEAKSDTGQTKVGRFOVTTTA 1864
2214 -----VLGGGEGDIE-----PVSPPEGMTPEGHSRAVYPLLRYDGEQTPPSRMGSKS 2261
1865 NKVGRFSVSKTEDKITDTKKEGFPVASPPFMDLEQAVLPAPV PKKEKP-ELSEPSHLN---1920
2262 PGNTSQP-PAPFSKLTESAMVKSQKQEKINKLNTNREPEYNIQSPTGTFINMPAIT 2320
1921 -GPSSDPEAAFLSRDVGSGSPHSPHQLSSKSL-----PSQNLSSLSNSFNSSYMS 1972
2321 GTGLMTYRSQAVQ-----EHAGTNMGLEAIIRKALMGKYDOWEESPPLSANAFNPLN 2372
1973 SNESDIEDLKLRLRLRDKHLKEIQDLSQRKHIESLYTLKGVPP-----2022
2373 ASASLPAAMPITAAOGRSDHTLTPSGGGKAKVSGRPSRKAKSP-APGLASGDRPPSV- 2430
2023 -AVIIPPAAPLS---GRRRRPTKS-----KGSKSRSSSLGNKSPQLSGNLSCQSAASVL 2073
2431 ---SSVHSEGD 2438
2074 HPQOTLHPGN 2084

RESULT 77
ABO44390
ID ABO44390 standard; protein; 2217 AA.
XX ABO44390;
XX AC ABO44390;
XX DT 26-SEP-2003 (first entry)
XX DE Novel human protein kinase #10.
XX KW Human; kinase; enzyme; cosmetic application; nutraceutical application.
XX OS Homo sapiens.
XX PN US6541252-B1.
XX PD 01-APR-2003.
XX PF 14-MAY-2001; 2001US-00854856.
XX PR 19-MAY-2000; 2000US-020601SP.
XX PA (LEXI-) LEXICON GENETICS INC.
XX PI Walke DW, Hilbun E, Donoho G, Turner CA;
XX WPI; 2003-575927/54.
XX N-PSDB; ACH03773.
XX PT New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.
XX PS Disclosure; Page; lipp; English.
XX CC The invention relates to a new isolated nucleic acid encoding a novel

CC human protein kinase. The nucleic acid is useful in cosmetic and
CC nutraceutical applications. The present sequence represents the amino
CC acid sequence of a novel human protein kinase. Note: The sequence data
CC for this patent did not from part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6541252B1
XX
SQ Sequence 2217 AA;

Query Match 3.2%; Score 422; DB 7; Length 2217;
Best Local Similarity 19.8%; Pred. No. 1.8e-14;
Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;

QY 147 PVSPPSPHTDPELELAVPRLKEELIQNMVDRDREITVVEQISLKKKQOQLEBAAK 206
DB 99 PLSLPQPSIPAAPQAPPEPHRETV-----TATATSOVAQAPPAAPGEQAVA 149
QY 207 PPEP-----EKVSPSPPTESKHSRLVQIIYDENRKAABAHRIELGLQFQVPELPLY 257
DB 150 GPAPSTVSPSTSKDRPVSPSL-----VGSKEPPPPA 181
QY 258 NQPSDTRQYHENIKINQAMRKKLILYKERNHARKQKFCQRYDOLMEALEKKVERIE 317
DB 182 RSGSG-----GGSKEPQBERSQOQDDI-EELETKAVGMS 215
QY 318 NNPRRRAKESKV-REYIEKQFPEIRKOR-----ELQERMOSRVQORGSLGMSAARGE 369
DB 216 NDRGRFLKFDIEICRGSKFYKGLDTTVEVAVCELODRK-----LTKSERQRF 265
QY 370 HEVSEIIDGLSGOENLEKQMRQAVIPMLYADQORIK-----FINNGML-ADPMKVYK 424
DB 266 KBEAEMKGL-OhpNIVR-----FYDSWESTVAKKICVLVTELMTSGTLKTYL 313
QY 425 DR-----QVMNWSQEKETPREKFWQHPK-----NFGLIA 455
DB 314 KRPFKWKIKVLSWCQILKGLQFLHTRTPPIIHRDLKCDNIFITGTSVKIGDLGL-- 371
QY 456 SFLERTVAECVL-----YYLTKNENY-----KSLVRRSYRRRGKSQOQOQOQ 501
DB 372 ATLKRASPAKSVLGTPEFWAPMEYKEYDSDVDYAFGCMLEMAISEYVSECQNAQI 431
QY 502 QOQOQOQOQO-----MPRSSQ-----EKKDEKEKEAEBKEKEKEPEVENDKEDLLK 547
DB 432 YRRVTSGVKSPASDKVAIPVKEIIEGCIRQNKDERYSIKOLLNHAFFQOETGVRVELAE 491
QY 548 EKTDDTSGEDNDEKAVASKRTANSQGRKRITSMANEANSE-EALTQOQSAELAS 606
DB 492 E-----DPGEKIAIKLWRIEDIKKLGKYGKDNKNEAIEFSLDERDVPEDVAQ--- 538
QY 607 MELNERRTEEMETAKKGLLEHGRNWSAIRMVSGKTVSQCKNFYFNYKKRQNLDEIL 666
DB 539 -EMVESGYCEGHDKTMAKAIDR-----VSLIK-----RKREORQLVR 576
QY 667 QQHLKMEKERNARRKKKA-----PAAASEEAPPPVPEDEMEASG 709
DB 577 EEQKKKQOESSLKQOQVEQSSASQGIKOLPSASTGPTASTTSASVTSQVEPEEPA-- 634
QY 710 VSGNEEMVEEAEALHASGNEVPRGCSGPATVN--NSSDTEISIPSPHTEAAKDTQNGP 767
DB 635 -DQHQOLQYQOQPSISVLSGTVDSGGQSSVFTESRVSSQOQTVSYGQCH-EQAHSTGTVPG 692
QY 768 KPPATLGADGPPPG--PPT-----PPRTSRAPTEPT-PASEATGAPTPPPAPPS 814
DB 693 HTPSTVQAOQSPGVVPPPSVQOQIQOTAPQQTQVQYSISQTSSEATTA-----QPVS 747
QY 815 PSAPPVVPVKEEBEETAAAPVE--EGEQKPPAAEE--LAYDTGKAEPPVKSECTEBA 870
DB 748 QPQAPQVLPQVSAKGLPVSQPVPTIQGEPIQIVATQPSVVPVHSGAHFLPV----- 799
QY 871 EEPGAKGDAEAAEATAEGALKAEKGGSGRATTAKSSCAPQDSSSATCSADEYDEAE 930
DB 800 -----GQPLTPL-----LPQYVPSQIPISTPHVSTAQ 827

QY 931 GGDKNRLLSRPSRLTLPTGCDPRANASPOKPLDLKQLKQRAAIP-----PIQTVKHBP 984
DB 828 TG-----FSSLPTWA-----AGITQPLTLTASSATTAAPGVSTVVPSPQLTLOP 874
QY 985 PREDAATKPAAPPPPPONLOPESDAPQOQSSPRGKSRSPAPPADKAEFAAEAKLPG 1044
DB 875 VTQ-----LPSQVHPQLQP-----AVQSMGI-----PAVL-GQAAEVLSSG 911
QY 1045 DPCWTSGLFPVPVPREVIKASPHAPDPSAFYAPPGHPLPLGLHDTARPVL--PRPPTI 1102
DB 912 D--VLYQGFPPRLPQY-----PGDSNIAPSNNVASVCIHST--VLXPPMPTEV 956
QY 1103 SNPPPLISAKHPV--LEROIGALSQMSVOLHVYPYSEHAKAPGVPTMGLPLPMDPK 1159
DB 957 LATPGVFTVQVYVESNLLVPMGV--GGQVQVSPGGSQAQAPTSSQQA-- 1007
QY 1160 KLAPSGVKQEQLSRPGQAGBPESLGV-----PTAQEASVLRTGALSGVPGSITKGI 1212
DB 1008 -LESTQGV-----SQVAPAEPVAVAQOQTPTTLASSV--DSAHSDVAG-MSDG- 1054
QY 1213 PSTRVPSDAITYRGSITHGTADVLTKGTITRIIGEDSPSRDLDRGREDSLPKGHVIEG 1272
DB 1055 -NENVPSSG-----RHEGRTTKHYRKSVRSRSEKTSRPKLRILNVS 1098
QY 1273 KGHVLSYEGGMSVTQCSKB-----DGRSSGPPHETAAPKRTYDMMGRVG 1319
DB 1099 NKGD-----RVVEQCLETHNRKMVTFKDLGDN-----PBEIATIMVNDFI----- 1141
QY 1320 RAISSASIEGLMGRALPPEHSPHILKEQHHRIGSITQIGIPRSYVBAQEDYLREAKLLK 1379
DB 1142 LAIERESFVDQVREII--EKADEMLSEDSVVEPEGDQGL--ESLQKDDYDYGSGSKLE 1196
QY 1380 REGTTPPPPPSRDLTEAVYKTOALGPLKLPKPAHEGLVATVKEAGRS--IHEIPREERLHTP 1437
DB 1197 GEFKQPIPASSM-----PQQIGIPSSLTQVHVSAGRRRIVPVPBSRLRESK 1244
QY 1438 ELPAPRLUKESITQGTPLKYDTGASTGSK-----KHDVRSL----- 1476
DB 1245 VFP-----DTVAATAQSPGMNLSHASSLSLQOAFSELRAQMT 1289
QY 1477 -----IGSPCRFPVPHPLDVMDARALERACYEESILKSRPGTASSSGSIARGAP- 1527
DB 1290 GNTAPPNFSHTGPTFPVVPP-----FLSIAAGVTTAAATAPVATSSPP 1335
QY 1528 -----VIVPELCKPRQSL--TYEDHGAPFAGHLPGRGSPVTMREPTRLQEGSLSSKA 1579
DB 1336 NDISTSVIOSEVTVPEEGIAGVATSTGVVTSGL-----PIPPVSESVLSSVV 1385
QY 1580 SQDRKLTSPTRIA--KSPHSTVPBHHPHIPISPYEHLRGVSGVDLYRSHIPLAFDPTS 1636
DB 1386 SS-----ITIPAVVVISITTSFSLQVPTSTSEIV-----VSSITALYPS----- 1422
QY 1637 IPRGIPLDAAYAYLPHRLAPNPTYPHLYPPYLI-----RGVYDPT 1676
DB 1423 -----VTVSATSASAGSGTATPGK-----PPAVVSOQAAGSTTVGATLTSVTTSPST 1473
QY 1677 ALENRQ-----TIINDYITSQOMHNTATAMA-----QRA 1707
DB 1474 ASQLSLSSTSTPTLAETVVVSAHSLDKTSHSSTGLAFSLASAPSSSSPGAGVSSYI 1533
QY 1708 DMLRGLSPR-----ESSIALNYAAGRGIIDLSQVPHPLVPLVPTFGTATAMRLAYL 1761
DB 1534 SQPGLHLPLVPSVIASTPILQOAGPTSTPLLPQVPSIPPLVQPVANVPAV---QOTLI 1590
QY 1762 PTAPQPFSSRSHSSPLSPGCP--THLTKPTTSSSERERDRDRDREREKESILSTTT 1820
DB 1591 HSQOPQ-----ALLPNQPHTCP-----EVSDDTOPKAGIDDIKT 1626
QY 1821 VEHAPTRPCTEQS GSGSGSGSGSGSGSRPASHASHAHQHSPISPRTOALQORPSVLHN 1880
DB 1627 LEE-----KLRLSEHSSSGA-----QHASVSLTS-----LVIES 1658
QY 1881 TGMKGI-ITAVEPSKPTVLRSTSTSTSFVRPAATFPFATHCPLGLGLDGVYPTLMEPVLLP 1939

Db 688 QQAQOVVLQVSGAGKQLPVSQPVPTTQGEPOIPVATOPSPVVPVHGAHFLPV----- 739
QY 871 BEGPAKGDAEAAEATAEGALKAEKKEGSGRATTAKSSGAPQDSDSSATSCADEYDEAE 930
Db 740 -----GQPLPTPL-----LQYVPVQPIPISTPHVSTAQ 767
QY 931 GGDKNLLSPRSLTPTGDPANASPOKPLDLKQLKQRAAIP-----PQVTKVHEP 984
Db 768 TG-----FSSLPTMA-----AGITQPLTLASSATTAIPGVSTVVPVQPLTLQP 814
QY 985 PREDAAPTKAPAPPAPPPQNLQSPESDAPQPGSSPRGKSRSPAPPADKEAFAAEAKLP 1044
Db 815 VTQ-----LPSQVHPOLLQP-----AVQSMGI-----PANL-GQAAEVLSSG 851
QY 1045 DPPCWTSGLPFPVPPREVIKASPHADPPSAFSAYPEGHPHLPGLHDTARPVL---PRPPTI 1102
Db 852 D---VLXQGFPPRLPPQY-----PGDSNIAPSSNVAVCIHST---VLXPPMPTEV 896
QY 1103 SNPPPLISSAKHPSV---LERQIGALISQMSVOLHVPYSEHAKAPVPTMGLPLPMDPK 1159
Db 897 LATPGYFPTVQPVYBSNLLVPMGGV---GGQVQVSQFGGSLAQAPTSSQQA----- 947
QY 1160 KLAPFSGVQEQLSPRGAGPPESLGV-----PTAQASVLRGALGSPVGGSTTKGI 1212
Db 948 -LESTQGV-----SQVAPAEVVAQPOATQPTTLASSV---DSAHSDVASG-MSDG- 994
QY 1213 PSTRVPSDAITYRGSITHGTADVLVYKGTITRIIGEDSPSRILDRGREDSLPKGHVYIEG 1272
Db 995 -NENVESSG-----RHGFTTKRHYKSVRSRSRHEKTSRPLKRLNVS 1038
QY 1273 KXGHVLSYEGGMSVTCQSK-----DGRSSGPPHETAAAPKRTYDMEGRVG 1319
Db 1039 NKGD-----RVBECQLETHNRKMTVKFDLDGDN---PEEATIMVNNDFI----- 1081
QY 1320 RAISSASIEGLMCRAIPPERHSPHLLKEQHIRGSITQGIIPRSYVEAQEDYLREAKLJK 1379
Db 1082 LAIERESFVDQREII---EKADEMLSEDSVBEPEGDQGL---ESLQKDDYFGSGSKLE 1136
QY 1380 REGTPPPPPSRDLTEAYKTQALGPLKLPKPAHGLVATVKEAGRS---IHIEPREELRHTP 1437
Db 1137 GFQKQIPASSM-----PQIGTPTSSLTQVHVSAGRRIVPVPSRLRESK 1184
QY 1438 EULPAPRLKEGSIQTGTLKYDTGASTTGSK-----KHVRLS----- 1476
Db 1185 VFP-----DTVAASTAQSPGMNLSHSSLSLQQAFAFSELRRQAMTE 1229
QY 1477 -----ICSPGRTFPVPHPLDMADARALERACYEESLKRPGTASSSGGSIARGAP- 1527
Db 1230 GNPATPNTSHTGPTFPVPPP-----FLSSITAGVPTTAAATAPVPATSSPP 1275
QY 1528 -----VIVPELKGPRQSPPL---TYEDHGAPFAGHLPKRGSPVTMREPTPRLEQSGLSSSKA 1579
Db 1276 NDISTSVIQSEVTVPEEGTAGVATGTGVTSGL-----PIPPVSESPVLSVV 1325
QY 1580 SDRKLTSTPREIA---KSPHSTVPEHHPIPSIPEYHLLRGVSGVDLYRSHIPLAFDPTS 1636
Db 1326 SS---ITIPAVVSIITSPSLQVPTSTSEIV-----VSSTALYPS----- 1362
QY 1637 IPRGIPLDAAYYLPHRLAPNTYTHLYPPYLI-----RGYPDT 1676
Db 1363 -----VTVSATSSAGSGSTATPGPK-----PPAVVSOQAAGSTTVGATLTSVSTTSPFST 1413
QY 1677 AALNRQ-----TIINDYITSQMHNTATAMA-----QRA 1707
Db 1414 ASQLSIQLSSTSTPTLAETVVVVAHSILDKTSHSTTGLAFSLASPSSSSPGAGVSSYI 1473
QY 1708 DMLRGLSPR-----ESSALNVAAGRGIHIDLSQVPHLPVLVPPPTGPTATMDRLAYL 1761
Db 1474 SQPGGLHPLVIPSVIASTPILQAAGTSTPLLPQVPSIPLVQPVANVPAV---QOTLI 1530
QY 1762 PTAQPFSSHSSPLSPGCP---THLTKPTTSSSERDRDRDREREKILSTTTT 1820
Db 1531 HSQFQP-----ALLPNQPHTHCP-----EVDSDTOPKAPGIDDIKT 1566

QY 1821 VEHAPIWRPGTQSSGSSGSSGSSGSSSRPASHSHAHQHSFISIPRTQDALQORPSVLHN 1880
Db 1567 LEE-----KLRSLSFSEHSSGA-----QHASVSLETS-----LVIES 1598
QY 1881 TCMKGI-ITAVBPSKPTVLRSTSTSPVRAATFPDPATHCPGLGTLGDGVVPTLMEPVLLP 1939
Db 1599 TVTPGLIPTTAVAPSK---LTSIT---STCLPTNLPLG----- 1631
QY 1940 KEAPRVARPERADTGHAFKAPPARSGLEPASSPSKSEPRPLVPVSGHATIAITPA 1999
Db 1632 ---IVALPVTVPVTPGVQVSTPVSITTSVGKGTAPSKPPLTKAPVLPVGTETLPACTLPS 1687
QY 2000 KNLAPHASDPDPAPPASASDPHREKTS- KFSIQEELRSLGYSYSSSPGVPFVSP 2058
Db 1688 EQL-----PPFPGPSL-----TQSQOPLDLDQAQR-----RTLSPXITVTS 1726
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGELRPQGPVKLGEEAHLPHLRPLPS---SQPS 2115
Db 1727 V-----GPVSMAPTA-ITEAGTOPKGVSOVK 1753
QY 2116 SSPLQLQTAGV---KGHQRVVTLAQHISEVITQDYTRHHPQQLSAPLPAPLYSFFCASC 2171
Db 1754 EGPVLATSSGAGVFKMGREFQSVAA-----DGAQKEGKNKSEDAKSVHESSTSES 1804
QY 2172 PVLDRRPPSDLYLPPP-----DHGAPARGSPHSEGGKRSPEPNKTS----- 2213
Db 1805 SVLSSSSPESTLVKPEPNGITIPGISSDVPESAHTTASEAKSDTGQTKVGRFQVTTTA 1864
QY 2214 -----VLGGGEGDIE-----PVSPEGMTPEGHSRSVYPLLYRDGQTEPSRMGSKS 2261
Db 1865 NKVGRFSVSKTEDKTDTKKEGVASPPFMDLEQAVLPVPIPKKRP-ELUSEPSHLN--- 1920
QY 2262 PGNTSOP-PAFTSKLTESNAMYKSKQIKKLNTHRNEPEYNISSQGTETIFNNPAIT 2320
Db 1921 -GPSSDPEAAFLSRDVGSGSPHSPHQLSSKSL-----PSQNLSQLSNSFNSSYMS 1972
QY 2321 GTGLMYRSQAVQ-----BHASTNMGLEAIIRKALMGKYDQWEEPSPLSANAFNPLN 2372
Db 1973 SDNESDIEDLKLRLRLDKHLKEIQDLQSRQKHEIESLYTKLGKVP- 2022
QY 2373 ANASLPAAMPITAADGSDHLLTSPCGGKAKVSGRPSRKAQSP-APGLASGDRPPSV- 2430
Db 2023 -AVIIPPAAPLS---GRRRRPTKS-----KGSKSRSSSLGNKSPQLSGNLQQAASVL 2073
QY 2431 ---SSVHSEGD 2438
Db 2074 HPQQTLLHPPGN 2084

RESULT 79

ABO44389
ID ABO44389 standard; protein; 2354 AA.

XX ABO44389;

XX 26-SEP-2003 (first entry)

XX Novel human protein kinase #9.

XX Human; kinase; enzyme; cosmetic application; nutraceutical application.

XX Homo sapiens.

XX US6541252-B1.

XX 01-APR-2003.

XX 14-MAY-2001; 2001US-00854856.

XX 19-MAY-2000; 2000US-0206015P.

XX (LEXI-) LEXICON GENETICS INC.

XX	Walke DW, Hilbun E, Donoho G, Turner CA;		
PI	WPI: 2003-575927/54.		
XX	N-PSDB; ACH03772.		
DR	New nucleic acid encoding novel human proteins, useful in cosmetic and		
XX	nutriceutical applications.		
PT	Disclosure; Page: 11pp; English.		
PS			
XX	The invention relates to a new isolated nucleic acid encoding a novel		
CC	human protein kinase. The nucleic acid is useful in cosmetic and		
CC	nutriceutical applications. The present sequence represents the amino		
CC	acid sequence of a novel human protein kinase. Note: The sequence data		
CC	for this patent did not from part of the printed specification but was		
CC	obtained in electronic format directly from USPTO at		
CC	seqdata.uspto.gov/sequence.html?DocID=6541252B1		
XX			
SQ	Sequence 2354 AA;		
Query Match 3.2%; Score 422; DB 7; Length 2354;			
Best Local Similarity 19.8%; Pred. No. 2e-14;			
Matches 513; Conservative 291; Mismatches 943; Indels 844; Gaps 114;			
QY	147 PVSPPPPHPTDPELELVPPRLSKEELIQNMDKVDREITWVEQIQISKLKKKQOQLEEEAAK 206	QY	768 KPPATLGADGPPPG--PPT-----PPRRTSRAPTEPT-PASEATGAPTPPPAPPS 814
DB	99 PLSLPQPSIPAAPVQSAPEPREHEIV-----TATATQVAQPPAAAAPGEQAVA 149	DB	693 HIPSTVQAQSPHGVYPPSSVQOQIQOTAPQOTVQVLSQTSSTSEATTA-----QPVS 747
QY	207 PPEP-----EKPVSPPPIESKHSRLVQIYDENRKKAAHRLGLEGPQVVELPLY 257	QY	815 PSAPPPVVPVKEKEBEETAAAPVE--EGEEQKPPAAEE--LAVDTGKAEEPPVKSECTEEA 870
DB	150 GPAPSTVPSSTKDRVSPSL-----VGSKEEPPPA 181	DB	748 QPAPQVLPVQSAGKQLPVSQVPTIQGEPIPVATQPSVVPVHSGAHLFPV-----799
QY	258 NPSPDTRQVHENIKINQAMRKKLILYFKRNHARKWKQKFCQRYDQDLMEALEKKVERIE 317	QY	871 BEGPAKGRDAEAAEATAEGALKAEKKEGSGRATTAKSGAPQDSDSSATCSADEVDEAE 930
DB	182 RSGSG-----GSAKEPQEERSQOQDDI--EELETKAVGMS 215	DB	800 -----GQPLPTEL-----LPQVPSQIPISPHVSTAQ 827
QY	318 NNPRRRAKESKV-REYEFKQFPEIRKQ-----ELQRMQSRVQSGSLSMSAARSE 369	QY	931 GGDKNRLSPRSLTLPTCDPRANASPOKPLDLKQKQRAAIP-----PQTQVTKVHEP 984
DB	216 NDGRFLKFDIEICRGSKFYVYKGLDTETVEVAVMCELDQRK-----LTKERQRF 265	DB	828 TG-----FSSLPIWA-----AGIQPLTLTASSATTAAIPGVSTVVPVSQLPTLLQP 874
QY	370 HEVSEIIDGLSGQENLEKQMLQVLAIPMLYDADQRIK-----FINNGLM-ADPMKVYK 424	QY	985 PREDAAPTKAPAPPQPNQLOPESDAPQOQSGSPRGKRSRSPAPPADKFAFAAEAKQKUPG 1044
DB	266 KEEAEMLKGL-QHPNIVR-----FYDSWESTVAGKKCVILVLTMLTSGTLKTYL 313	DB	875 VTQ-----LPSQVHPQLQLP-----AVQSMGI-----PANL-GQAAEVPPLSSG 911
QY	425 DR-----OVNMWMSQEKETREKFMQHPK-----NFGJLA 455	QY	1045 DPCCWTSGLPPFPVPPREVIVKASPHAPDPSAFSAVAPPGHPLPLGLHDTARPVL--PRPPTI 1102
DB	314 KFKVMKIKVLRSWCKRQILKGLQFLHTRTPPIIHRDLKCDNIFITGPTGSVKIGDLGL-- 371	DB	912 D--VLYQGFPPRLPQY-----PGDSNIAFSSNVASVCIHST---VLXPPMPTEV 956
QY	456 SFLEKRTVAECVL-----YYILTQKNENY-----KSLVRSYRRRGKSGQQOQOQ 501	QY	1103 SNPPPLIISAKHPSV---LEROIGAIQSMVSQVQLHVPYSEHAKAPVGPVTMGLPLPMDPK 1159
DB	372 ATLKRASFVAKSVIGTPEFWAPMEYEEKYDESVDVYAFGCMLEMATSEVPYSECQNAQI 431	DB	957 LATPGYFTVTVQVYVESNLLVPMGGV--GGQVQVSPQGGSLAQAPTSSQAV-----1007
QY	502 QOQOQOQOQO-----MPRSSQ-----EKKDEKEKEAEKKEEKEVEVNDKEDLLK 547	QY	1160 KLAFFSGVQKQEBOLSPRGAGPPESLGV-----PTAQEASVLRGTALGSPVGGSIITKI 1212
DB	432 YRRVTSGVPASFDKVAIPEVKEIIEGCIQKQNDERYSIKOLLNHAFFOEETGVRVLAEL 491	DB	1008 -LESTQGV-----SQVAPAEVAVAQOATQPTTLASSV--DSAHSDVASG-MSDG- 1054
QY	548 EXTDDTSGEDNDEKAVASGRKRTANSQGRKGRITRSMANEANGSE-EAITPQQQAEALAS 606	QY	1213 PSTRVPSDAITYRGSITHTGPADVLVYKGTITRIIGEDSPSLDRGREDSLPKGHVIEG 1272
DB	492 E-----DQGEKIAIKMLRIEDIKKLGKGYKQNEAIEFSDFLERDVPEDVAQ--- 538	DB	1055 -NENVPSSG-----RHEGRTTKRHYKRSVRSRSHKETSREPLKRLINVS 1098
QY	607 MELNESSRWTEEMETAKGLLEHGRNWSAIRMVGSKTVSQCKNFYFNYKQNLNDEIL 666	QY	1273 KKGHVLSYEGGMSVTQCSKE-----DCRSSGPPHETAAPKRTYDMMEGRVG 1319
DB	539 -EWESGYCEGDHKTMAIKDR-----VSLIK-----RKREQQLVR 576	DB	1099 NKGD-----RVVECCOLETHNRKQWTFKPDLDGDN---PEEITATIMVNDFI-----1141
QY	667 QHKULMEKERNARKKKA-----PAAASEEAAFPVVEDEEWEASG 709	QY	1320 RAISSASTIEGLMGRAPIPERHSPHLKEQHHRIGSITQIGIPRASYVEAEDYLRRAKLLK 1379
DB	577 EEQKKKQOESSLKQOQVEQSASQTGKQLPQASGTGPTASTTSASVSTQVEPEPEEA--- 634	DB	1142 LAIERESFVDQVREII---EKADEMLSEDVSVEPEGDQGL--ESLQKDDYDFSGSQKLE 1196
QY	710 VSGNEEMVEEAELHASGNEVPRGCSGPAVFN--NSSDTEIIPSPHTEAAKTQNGQP 767	QY	1380 REGTPEPPPPSRDLTEAVKTAQALGPLKPAHEGLVATVKEAGRS--THEIPRELRHTP 1437
DB	635 -DQHQQLQYQPSISVLSDGTVDGQSGSVFTESRSSQOTVSYGSOH--EQAHSTGTVP 692	DB	1197 GEFKQPIFASM-----PQIGITPSSLTQVHSGARRFIVSPVPESLRRESK 1244

CC aureus. (i) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-
CC AB237516)

XX SQ Sequence 19938 AA;

Query Match 3.2%; Score 417.5; DB 6; Length 19938;
Best Local Similarity 20.5%; Pred. No. 5.9e-13;
Matches 476; Conservative 158; Mismatches 813; Indels 879; Gaps 114;
603 ELASMEINSSRWTEEMETAKGLLEHGRNWSAIARMVSKTVSOCKNFYFNKKQNL 662
1420 ELGQHELOAAA-EESPVDHAGQGDAGRPQGGGLHA----- 1456
663 DETLOQHLKMEKERNARKKKAPAAAEAAFPVVEDEEMEAAGVSGNBEEMVEAE 722
1457 -EPFGSGFRRGDRHVDRTGEAAPARTPGAGRTPRIEGHV---GCKGRE----- 1504
723 ALHAGNEVP--RGECSG-----PATVNNSSDTEISPSHTE----- 757
1505 --HAGDRPVAVRGAVHAGAGGHLFVLVRQRRPARPAPDPLRPRGPAPARTAVRRGRA 1562
758 -----AAKDTQNGKPKPA-----TLGADGPPPGPTTPPRTSRAPIEPT 797
1563 GSQDGGSHRTPGDPAPQVPADDELHRVGAGAVHRTGAES-PPRPAPPTRVRRTEGAR 1621
798 PAS-----EATGAPTPPAPSPAPVPPVPEKEEBETAAPVVEGEQKPPA--- 847
1622 PVSTHAHLRTRTGRPLHPPG--DARSPAPGLHREPTWPGFWP---GRGVHPGAQVC 1675
848 -ABELAVDTGKAEEPVK-----SECTERA----- 870
1676 DRAVTGHDAGVEEPARPSGGRXWSTALEWFKSSYSGSGQCVEALCPHTIHIRDSKN 1735
871 --BEGPAKGDABA---AEATAGALKAB-----KKEGSGR-----ATTA 906
1736 TPEDGTLQVPTAWAFAFTSATTEXARRPELTPTPTVPVPPRCGPGRCRRRPPVSTR 1795
907 KSSGAPQDSDSSATCSADEVDEAE-----GQDKN----- 935
1796 RPRCGKRRKTGTCTPASAGVVRPPLPCAGAGPSPGPGCGHGRPPGVRRSAXAAA 1855
936 RLISP--RPSLTPGTDPANASPOKPLDLKQLKORAA-----AIP- 974
1856 RRGSPRRPACXGFLRPSARAVRDRARRRRERRAAVGGRAACRPSARTHARFAVPA 1915
975 -----PIQVTKVHEPPREDAAPTKPA-----PPAP-----P 1000
1916 PCAAPSAPGAPVCPVRRRSGP-----GPACPAWLFPFPPGPAHPLWLAPLRTPARPAGLLP 1971
1001 P----PQNLOPESDAQOQGS--SPRCKSRSPADKAEAPAEAKLPG--DPPCWTSLG 1053
1972 FLARPPXAPRVLPRVPGTRTPRPSARSRRPPQRSRAAGAPSTAPDAETPC----- 2026
1054 PFVPVPREVIKASPHAPDPSAFSAPP-----GHPPLGLHDTARPVL----- 1096
2027 ---ASPSCEARAPRGPFPFPCWAPTRPCAGAPGVXGRPAGCADRPVPPCCTPWS 2083
1097 -PRPPTISNPPPLISSAKHPSVLEROIGAISQMSVOLHVFPY---SEHAKAPVGPVTMGL 1152
2084 PPRRRRRWPPPARAGSGAPG---CGTPRRGS-----PWYRSGTARGSAPPTRRR 2132
1153 PL-----PMDPKLAPFSGVKQEQLSPRQAGPSPESLGVPTAQ-EA 1192
2133 FMSGGSDSSRSRPPRRAPCPRRPR--APGSPAARRWXPRRRPPAPSSRRTRPAGRGA 2190
1193 SVLRGTALGS-VPGSGITTKIPSTRVPSDSAITYRGSITHGTADVLVYKGTITRIICEDS 1251
2191 AVSRAAHCAHSRPSGSXARWPDPPYXSGSAPRRLGARVRWP-----GSSCR-----A 2239
1252 PSRLDRGRDLSLPGHVIYEGKKGHVLSVEGGMVST---QCSKEDGRSSGPPHETAAPK 1308

2240 PBR-----RRGSGAG-----RGRSG--CRPPGSRVVRPGACRRAPGSGXPARPPRPACTPR 2289
1309 RTYDMWEG-----RVGRAISSASIEGLMGRAIPPERHSPHLLKEQHHRIGSITQIGIPRSYV 1364
2290 PPRGRGRGPARRAGRSTERSAARSPPGR--PP--RSP-----GGAAGRPRPGR 2334
1365 EAQEDYLRREAKLLKREGTPPPPPP--PSRDLUTEAYKTOALGPLKLPKPAHEGLVATVKEA 1421
2335 AAVAHRRGRPARSVR--GSPRRPPAPVVRPAPRTACAAAGLPPAPPRA-----A 2382
1422 GRSIHEIPREELRHTEPLPLAPRLKEGSIITQGTPLKYDTGASTTCKSKKHDRSLIGSPG 1481
2383 GRASSAPRPPRR--PGVPRGPPP-----PAR-----AARRPVR---WSCA 2418
1482 RTFPFVHPLDMADARALERACYEBSLKS-----RPGTASSSGSISARGAPV 1528
2419 RACP-----RRCVRECCVQPRRSAGRRPPSGRERSAARPCGAGAAGTS-RRRAP- 2466
1529 IVPELGKP--ROSPLTYEDHGAPFAGHLPGRSGPVTMRREPTPRLOBSLSSSKASQDRKL 1585
2467 -----GRPSGTRPSP-----PPGCAACPRXGPTAGPPXPPARAGPVAHG----- 2506
1586 TSTPREIAKSPHSTVPEHHHPHPISPVEHLLRGSVGVLDLYRSHIPLAFDPTSIIPRGIPLDA 1645
2507 -SVPGGPVVR-PRRRVRDRRPPPPAPPREAARPGPP-----PPAHSAPSRPR----- 2549
1646 AAAYVLPRLHPLNPTVPHLYPPYLIRGY-----PDTAALENROTIIINDYITTSQOMHNTAT 1701
2550 -----PSHRVPASCHP---GGGRIDGWACAGCAPTLTIGRRANTCPLYVGSR-----G 2593
1702 AMAQADMLURGLSPRESSIALNVAAGPRGIIIDLSQVPHL--PVLVPPPTGTPTATMDRLA 1759
2594 ALGQ-ADRLK-----AETGLNGPVRSPWPAVAPDPTRRSRAA 2629
1760 YLPTAQPPSSRHSSPLSPGGTHLTKTPTSSSEBERRDRDRDR----- 1808
2630 SXP-----TTGPGRAGDRDGPGRGAXSRVXPVRRPPVPA 2662
1809 -BREKSIILTSTTVEHAPITWRPCTEQSGSSSGSSSGSSSRPASHSHAHQHSPIIS--- 1864
2663 WRRCRALLPRRAL---PFWKRGXGADSAAPPNLFCAGGLSRRSPSPWSA---SPTSXCR 2716
1865 --PRTODALQOQPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPVPRPAATFPFATHCPLG 1922
2717 CAPDRGDARAOXPR-----RRPTAAAGSRPPPARPPSPGARRAG 2760
1923 -----GTLGVVPTLMEPVLLPKEAPRV-----ARPEPRAD 1954
2761 SRRRRRRSGSGSPSGT--GRCPCWPRPV--PSRTARCGRNPARRSAPXCRSRRRPRAP 2816
1955 -TGHAFLAKPPARS-----GLEPASSPSKG 1978
2817 RTG-----PPTSFRTFPRSPWSQGSASASVREAPGRRARHLPRGDQORPADQPRG 2870
1979 SEP--RPLVPPV-----SGHATARTPAKNLAP 2004
2871 DDPGTRGALAEVVRPGVAPCGRQASAGAGVRCVPGCGDITLGTCTCGVRRSSGLRAL 2930
2005 HHA-----SPDPAPPASADPHREKTSQKPFQIQLSLSLGSHSSSYSEGEVFPVS 2057
2931 RAARWALSNSRRPPGGRARAOGTIXRQGLGRP-----RLRVL---GDCLPAQRVGP-- 2979
2058 PVSSPSLTHDKGLPKH--LEELDCKSHLEG-----ELRKPQGPVK-----LGGEA 2100
2980 -----OPGHLVORADGGPAEEFAGGPGVDLPVVPGRARGGGKLPDLGGHA 3027
2101 AHL-----PHLRPLPESQPSSSLLOTPAPGVKHQRVVTLAQHISEVITODYTRHPQ 2153
3028 AHLDEQIDEPHRRHGLAASDGDQPRNTCAGG--GQHRVRDVA-HVDVVAV----- 3075
2154 QLSAPLPAFLYFPFGACPVLDLRRP--PSDLYLPPPDHGCAPARGSPHSEGGKSRSEPNK 2211
3076 -----GAQLQLHRRREHVDPDRLGEPPGH-----PASQADRRRGARP-- 3112

Db 733 PGSPHMAQOH-----GDPATTANNDVLSQMMPDVSIOQTNNVPPHVQAMQG 780
QY 1539 ---SPLTYEDHG---APFAG-----HLPRGSPVTMRBP-TPRLQEGSLSS 1577
Db 781 NSASGNHFSHGMSFNAPFSGAPNGMQSCGNQPGFVNKDVTLTSPLLVNLQSDISAG 840
QY 1578 KASQDRKLTST-----PREIAKSPHS-TVPEHHPHIPISPYEH-LLRGVSGVDLYR 1625
Db 841 HFGVNNKQNTWANKPKKPKKNSQODLNTPTDTRPAGLEAEADQPLPGEQGISLDN 900
QY 1626 SHIPLAFDPTSTPRGI---PLDAAAAYLYP---RHLPANPTVEHLYP-PYLIR----- 1671
Db 901 SG-PKLPFNSRPPGVPSPQVEORPQQMPPLQMQRVAPPPQPPQPPQPPQPPQPP 959
QY 1672 -GYPDTAALENRQTIINDYITTSQMHNTATAMAQRAADMLRG-LSPRESLALNVAAGPR 1729
Db 960 PSQPOSQOQQOQQOQQMMMLMWOQPKSVRLPVSNVHPPRGFLNPDQRMPMQSGSVP 1019
QY 1730 GIIDLQVPHLPVLVPPTP-----GTPATAMDRLAYLPTAQO-PFSS----- 1770
Db 1020 VMVSLQG-----PASVPPSPDKQMPMPVNTPLGNSRKMVYQESPNQSSSPLAEMASLP 1075
QY 1771 --RHSSSPLSPGCPHTLTKPTTTSSEERDRDRDREREKSLTSTTTVEHAPIWR 1828
Db 1076 EASGSEAPSVPGSPNNMPSHV-----VLPQNQLMMTG---K 1109
QY 1829 PGTEQSSGSGSGSGSGSGSPASHAHQSP-ISPRTOALQORPSVLHNTGMKGII 1887
Db 1110 PGSPPLSATQCATPQPPVNSLPSSHCH---HPNVAAPTQTS---RPTKPNRSPRY 1163
QY 1888 TAVESKPTVLSTSTS-SFVRPAATPTPATHCPLGGTLGGVYPTLMPEVLLPKEAPRVA 1946
Db 1164 PQTNNRPPSPSTSLSPER-----LNASIAGLFP-----POIN 1199
QY 1947 RPEREADTGHAFLAPPARSGLEPASPSKSGSEPRPLVPVSGHATIAITPAKNLAPHH 2006
Db 1200 IFLPPRPNLRG-----DOQGLNPTTLKAIGQAPSNLTWNPSNFAT----- 1241
QY 2007 ASPDPAPPASADPHREKTSQKPFQIELESLGYHSGSVSPGVPSPVSPSLTH 2066
Db 1242 -----PQTHKLDVVVN-----SKQNSGATKASPSNS----- 1271
QY 2067 DKGLPKHLELDKSHLEGELRPQPGVKLGGEAAHPLHRLPESQPSSSPLLTQAPGV 2126
Db 1272 -----RRSSPGSSR-----KTTSPSGRQNSKAPKL----- 1296
QY 2127 KGHQVVTLAHLISEVITQDYTEHHQQLSAPLAPLYSPGACFPVLDLRPPSDLYLP 2186
Db 1297 -----TLASQTNAAQLQ-----NVELPRNVLVSP 1320
QY 2187 PPDHGAPARGS-PHSEGGKRSPEPNKTSVLGGG---EDGIEPVSP----- 2228
Db 1321 TPLANPVPFGSPNNSSG--LNPNSTSVVAAGVVEDNKESLNVQDSCQNSQSRKEQ 1378
QY 2229 ---EGMTEPGHRSAYVP--LLYRDGEQTEPSRMGSKSPGNTSQPPAPFSKLTESNAVY 2283
Db 1379 VNIELKAQVAQEVKVMVPEQSKQDQSPDNKLPs-----VEENKLV 1422
QY 2284 KSKQKQINKLN--TNRNEPEYNISQPG-TEI-FNMPALTGTGLM-----TYRSQAVQ 2333
Db 1423 SPAMREAPTSLSQLLDNSGAPNTIKPPGLTDLVTPPVVSGBDLKKASVITPQLDLS 1482
QY 2334 EHASTNMGIL-----EAIIRKALMGKYDQWES-----PPLSANAF 2368
Db 1483 KEPSNGLNPHSNELCSSLVHPSESVSNVAFSIPVMRSPVSSSSISTPLPPNQITVP 1542
QY 2369 ---NPLNASASLPAAMPITAAADGRSDHLLTSPGGGKAKVS-----GRP----- 2409
Db 1543 VTSNPITTSANTSAAALPHTLQSLAMSTVMTMPNAGSKVMVSEQSAQAQSNARQFITPVE 1602
QY 2410 -----SSRKAKSPAPGLA--SGDRPPSVSV--HSEGDGN-RRTPLTNRWED 2452
Db 1603 INSSIIQVMKGSQSPSTIPAAPLTNLSGLMPPSVAVVGLHPQNIKFSAPVP----- 1656

QY 2453 RESSAGSTFFPY-----NPLIMELQAGVMASPPPPGLPAGSGPLAGHHAWDEPKPL 2505
Db 1657 -ENALSSSPAPNIQTGRPLVLSRATPVQLPSP---PCTSSPVV-PSHPPVQOVKEL 1708

RESULT 83

ABO52986
ID ABO52986 standard; protein; 1791 AA.
XX
AC ABO52986;
XX
DT 09-OCT-2003 (first entry)
XX
DE Human putative spliceosome associated protein (SAP) #48.
KW Human; SAP; spliceosome associated protein; ribonucleoprotein;
KW RNP complex; RNA affinity substrate; RNP assembly sequence;
KW spliceosomal complex; hnRNP complex; mRNA export complex;
KW mRNA localisation complex; RNA editing complex; intron complex;
KW H complex; telomerase complex; fragile X protein complex;
KW reverse transcriptase complex; gene splicing complex.

XX Homo sapiens.

OS US2003068803-A1.

XX 10-APR-2003.

XX 14-JAN-2002; 2002US-00047991.

XX 12-JAN-2001; 2001US-0261521P.

XX (REED/) REED R.

XX (ZHOU/) ZHOU Z.

XX Reed R, Zhou Z;

XX WPI; 2003-540885/51.

XX Isolating ribonucleoprotein complex, by contacting RNA affinity substrate
XX having ribonucleoprotein assembly sequence and affinity tag, with protein
XX mixture, subjecting complex formed to chromatography, affinity selection.

XX Claim 24; Page; 39pp; English.

XX The invention relates to forming (M1) an isolated ribonucleoprotein (RNP)
XX complex (C), involves contacting an RNA affinity substrate (S) comprising
XX an RNP assembly sequence (AS) and an affinity tag, with a protein mixture
XX to permit formation of (C) on AS, subjecting (C) to chromatographic
XX separation, and subjecting (C) to affinity selection, where the affinity
XX tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli
XX maltose binding protein) binds to an affinity matrix. Also included are
XX an isolated spliceosome preparation (isolated by (M1)), a RNA comprising
XX an RNP complex binding site and at least one phage coat protein
XX recognition site, a nucleic acid encoding the RNA, and treating (M2) a
XX subject having a disorder associated with abnormal RNP complexes (by
XX obtaining a sample of cells from a subject, purifying RNP complexes from
XX the cells of the subject by (M1), determining the presence in the
XX purified RNP complexes of one or more proteins, and normalising the
XX amount of RNPs in the subject. (M1) is useful for forming an isolated RNP
XX complex selected from a spliceosomal complex (selected from E, A, B and C
XX complex), an hnRNP complex, an mRNA export complex, an mRNA localisation
XX complex, an RNA editing complex, an intron complex, or an H complex. (M1)
XX is useful in a diagnostic assay for determining whether a subject has
XX abnormal RNP complexes, (M2) is useful for treating a subject having a
XX disorder associated with abnormal RNP complexes. (M1) is useful for
XX forming an isolated RNP complex such as a telomerase complex, a fragile X
XX protein complex, a reverse transcriptase complex or a gene splicing
XX complex. The present sequence represents a putative novel human
XX spliceosome associated protein (SAP) isolated by the methods of the
XX invention. Note: The present sequence is not shown in the specification
XX but was obtained from Genbank or Swissprot using the information provided

XX	20-JUN-2003 (first entry)	
XX	Protein sequence #SEQ ID 837.	
XX	Multiprotein complex; eukaryote; drug target; diagnosis.	
XX	Saccharomyces cerevisiae.	
XX	EP1258494-A1.	
XX	20-NOV-2002.	
XX	20-DEC-2001; 2001EP-00130253.	
XX	15-MAY-2001; 2001EP-00111774.	
XX	(CELL-) CELLZONE AG.	
XX	Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;	
XX	Marzioch M, Schultz JD, Superti-Furga GD;	
XX	WPI: 2003-250078/25.	
DR	N-PSDB; ACC61028.	
XX	New isolated protein complexes useful for diagnosing a disease or	
PT	disorder, or as a target for an active agent of a pharmaceutical,	
PT	preferably a drug target in the treatment or prevention of disease or	
PT	disorder.	
XX	Disclosure; SEQ ID NO 837; 17pp + Sequence Listing; English.	
XX	The invention relates to multiprotein complexes from eukaryotes. Proteins	
CC	of the invention and DNA sequences encoding them are given in records	
CC	ABR52568-ABR53903 and ACC6010-ACC61944 respectively. The complexes are	
CC	obtainable by using a protein as a bait and isolating the set of proteins	
CC	which is attached thereto from cells. Such protein complexes may comprise	
CC	up to 30 distinct proteins. Protein complexes of the invention are useful	
CC	for diagnosing a disease or disorder, or as a target for an active agent	
CC	of a pharmaceutical, preferably a drug target in the treatment or	
CC	prevention of a disease or disorder. Note: The sequence data for this	
CC	patent is not represented in the printed specification, but is based on	
CC	sequence information supplied by the European Patent Office. The complete	
CC	document is available on CD-ROM	
XX		
XX	Sequence 1226 AA;	
XX		
XX	Query Match 3.1%; Score 415.5; DB 6; Length 1226;	
XX	Best Local Similarity 19.7%; Pred. NO. 2e-14;	
XX	Matches 236; Conservative 176; Mismatches 440; Indels 343; Gaps 44;	
Qy	21 PHSLSYPVQIARTDVGLEYQH-----HSRDYASHLSPGSIQIQP 63	
Db	124 PESWNTST-IHTHTDIGNSYSRKTMGRYNPQSTSTNTHFFSALSNAPFFVANGSS 182	
Qy	64 RRPSSLSEFPQG-----NERSQELHLRPESHSLYPGLGK----- 97	
Db	183 RRPSSMDYSDPVTNKLSTNNVSSVNNNSP--HSYSSRNKWSRGITSPRPFDNHNVM 240	
Qy	98 ---SEMFIESKRP-----RLLELPDPLLRSPILLATCQ-----PAGE 133	
Db	241 TTTSTNTNSIHOREPFWKANSNTILKSTHSSQSSPSLHTKTFHDANKLDKPEASVKVETPSK 300	
Qy	134 DLTKDRSLTGKLEP-----VSPSPSP-----HTDP-- 158	
Db	301 DETKAIASHDNNFPPRKSVKSPNAPLEPNTIKVGEEDALGKVEHKVHSGREITAKEHTPVK 360	
Qy	159 -----ELEVPPLRSKBEELIQNMDRVDRREITWVEQIQISKLKKKQOOLEEAAKPEP--- 210	
Db	361 MKEHDELEARAKKYSKINIDCKQDEIWTAKTVASAVEVSKESQKELTRSVERKESPEIR 420	
Qy	211 --EKVPSPPIESKHSRLVQIITVDENRKKASAAHRIEGL---GPQVELPLYNQSPDTRQ 265	

Db 780 ----TFSYRTPDFPS--GTNSSQSSPSSAANS--GHRSTLHGLAPKLGQR 831
QY 1035 FAABAOKLPDPPCWTGSLFPVPPREVINKASPHA--PPPSAFSYAPPGHPIPLGLHDTA 1092
Db 832 YRSGRKRSAGNIP-----LSPLARTPTPTQPTSPQRSPLLGHSLG 874
QY 1093 RPYLPR--PTTINBPPLISSAKHPSVLEHQICAI SQGMSVOLHVYPYSEHAKAPVCPVTM 1150
Db 875 NSKIAQAFSGKMSPTTIVR-----HIVRPKSAEPSPSL-- 909
QY 1151 GLPLPMDPKKLPFPKGQOLSPRGQAGPPESLGVPTAQEASVLRGTAIGSVPGGSITK 1210
Db 910 -LKRVOSEKLSFSYSGDKKHLCSR-----KHSLEV--TOEE--VQREQOREAPLOSIDE 960
QY 1211 GIPSTRVPSDAITYGSGTHGTPADVLYKGTITRIIG--EDSPSLRDGREDSLPKGHVI 1269
Db 961 NVCDVPPLSRAREVEQGCL-----KRPVSRKVGRQESVDDILDR--DKL--KAKVV 1006
QY 1270 YEGKGHVLSEYEGMSVTQCKEDGRSSSGPPHETAAPKRTYDMMGRVGRASSASIEG 1329
Db 1007 VK-----KADGF----- 1013
QY 1330 LMGRAIPPERHSHPHLKEOHIRGSGITGIPRSYEAQEDYLRRKALLKREGTTPPPPP 1389
Db 1014 -----PEKQESH--QKSHGPGSDLENFALFKLE-----EREKKVY-----PK 1048
QY 1390 SRDLTAYKTOA-----LGPLKPAHEGLVATVKEAGRSIHEIPRELHRTHELPLA 1442
Db 1049 AVERSSTFENKASQMBAPPLGSLLDKALHQASVRASEGAMSDGVPVPAEHRQGGDFRRA 1108
QY 1443 PRPLKEGSGITQGPLKYDGTGASTTG--SKKHVRSGLSGPRTTTPPVHPLDVMADARAL 1499
Db 1109 PAP--GTLQDGLCHSLDGIISGKBGTEKSKQAKELLRCE--KLDSKLANIDILYRKQMSL 1164
QY 1500 ERACYEESLKSRPGTASSGSGSIARGAPVIVPELKGPROSLTYEDHGAPFAGHLPRGSP 1559
Db 1165 EDK--EDNL-----CPVL-----KPKMTAGSHE-----CLP-GNP 1191
QY 1560 VT-----MRPTPLRQGS--LSSSKASQDRKLTSTPREI-----AKSPHSTVFEHHPH 1606
Db 1192 VRTGQQQPPPPASERAFVSSTHAAQMSAVSFVLKALTGRVDSGTEKPGILVAPE--SPV 1250
QY 1607 PISPVHEHLRG-----VSGVDLYRSHIP--LAFDPT-----SIPR 1639
Db 1251 RKSPEYKLEGRSVCLKPIEGTLDIALSGQASKTELPSPESAQSPSGDVASVPP 1310
QY 1640 GIPL-----DAAAYYL--PRHLAPNPTYPHLYPYLIRGYPDTAALENRQTI--NDY 1689
Db 1311 VLPSSSGKKNDDTSARELSFSSLLKNNKSY--LLEPWFL--PPSRGLQNSPAVSLPDPEF 1365
QY 1690 ITSQOMHNTA--TAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745
Db 1366 KDRKGPHPTARS PGVFMESNPGQREGSGPKHQ-----DHTTDPKLLTCLIGLNLHSPDLAR 1421
QY 1746 PTFGTATAMDLAVLPTAPQPFSSRHSSPLSPGPGTHLTKPTTTSSSER-----ER 1798
Db 1422 P-----RCPLPPEASPSREKPGLRRESSERGPTARSER 1454
QY 1799 DDDR-----ERDRDREREKIL-----TSTTTVEHAPIWRPGTQS 1834
Db 1455 SAARADTCHREPMELCFPETAKTSDNSKNLLSVGRTHPDFTYTQAMEKA--WAPG---- 1508
QY 1835 SGSSSGSGGGSSSRPASHSHAHQHSPISPRTQDALQORPSVLHNTGMKIITAVEPSK 1894
Db 1509 -GKTHHKGPG--EARPPPRDMSLHAGICEKELGKVR-----RQVERKPP 1552
QY 1895 PTVLRSTSTSPVRPAATPPATHCFLGGLTGVYPTLM-----BFLVLPKEAPVARPE 1949
Db 1553 EALLARRSLQ-----PPGIESEKSEKLSS--PFSLQKDGAKPE--RKEQPLQRHPS 1600
QY 1950 R-PRADTGHAFKAPAR-----SGLEPASSPS--KGSEPRPLVPVSGHATIR 1996

Db 1601 SIPPPLPTAKDLSSPAARQHCSSPSHASGREPGAKPSTAESSSPQDPPKVAHNS---- 1656
QY 1997 TPAKNLAPHASDPDPAPASADPHREKTKSQKPSFI-----OELELRSLG---YHGSSY 2048
Db 1657 ----ESSGHKPRPGDPGPKTKHDPDSLSSQKP--SVGATKGKEPATQSLGSSREGKCH 1711
QY 2049 SPFG--VEPVSPVS--SPSLTHDKGLPKHLEELD KSHLEGELRPKQGP--PVKLGGEAA 2101
Db 1712 SKSGPDVFPATPGSQNKASDGIQGEQGSFVPLHTDRAPLDAKQPTSGGRFLEVLKPV 1771
QY 2102 HLPHLRPLPESOPSSSPLLQTAPGVKGHQRVVTLLAQHISEVITQDYTRHHPOOLSAPLPA 2161
Db 1772 HLPFRCHPCSEPAQKL--SAVGEK-----QTLSPKHPK-----PS 1806
QY 2162 PLYSPFGASCPLDLR--RPPSDLYLPPDHGAPARGSPHSEGGKRSPE-----PNKTSV 2214
Db 1807 TVKDCP--TLCKQTDNRQTDKSPSQ-----PAANTDRRAEGKKCTEALYAPAEQDK 1855
QY 2215 LGGGEDGIEPVSPPEGMPTEPGHRSADVLYLLYRDGEQTEPSRMGSKSQNTSQPPAFPSK 2274
Db 1856 LEAGLSFVHSENRLLKGAERPAAGVGKGF--EARGKGGP--POKPP----- 1897
QY 2275 LTESNSAMVKSQKQINKKLANTHNREPEYNIISQPTGTEIFNMPAITGTGLMTYRQAOVE 2334
Db 1898 -TEAD-----KPNGMKESP-----SATGQSSFRSTALPE 1925
QY 2335 H-----ASTNMGLEAIRKALMGKYDQWESPPLSANAFNPLNASASLPAAMPITAADGRS 2390
Db 1926 KSLSCSSFPETRAGVREASASSD-----TSSAKAAGGMLLELPAPSNR 1969
QY 2391 DHTLTSPGGGK--AKVSGRPSR----- 2412
Db 1970 DHRKAQAGEGRTMTKSDSLPSFRVSTLPLESHHDPNTMGASHRDRALSVTATVGET 2029
QY 2413 KAKSPAGLASGRPPSVSSVHSEGDGNCNRRPTLTNRVWEDRPSS 2456
Db 2030 KGKDPAPA-----QPPARKQNVGRDVTKPSAPN--TDRPIS 2065
RESULT 87
ID AAM78959 standard; protein; 2161 AA.
XX AC AAM78959;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein SEQ ID NO 1621.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US004098.
XX PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX PA XX

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
XX N-PSDB; AAK52092.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX Claim 20, Page 3957-3961; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM7823-AAM80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication
XX Sequence 2161 AA;
XX
XX Query Match 3.1%; Score 414; DB 4; Length 2161;
XX Best Local Similarity 20.0%; Pred. No. 5e-14;
XX Matches 452; Conservative 245; Mismatches 787; Indels 780; Gaps 112;
XX
XX 471 YLTGKNEYKSLVRRSYRRGKSGQQQQQQQQQQQQQQQQQQQQQ-----KDEKEK 525
XX 373 YEVKQHFPSRLDWSLLR-----QKAEFIPQLSEDDTSYFDRSEK 415
XX
XX 526 EKEAEKEEKEPVENDKEDLLKKTDTDTSGED-NDEKEAVAS---KGRKTANSQGRKGR 581
XX
XX 416 YHHMETEE-----DQTNDEDFNVEIRQFSSCSHRFSKVFSS-----IDR 455
XX
XX 582 ITRSMANEANSEAITPQQAELASME-LNESRWTE-BEMETAKGLLEHGRNWSAIAIR 639
XX
XX 456 ITQNSAEE--KEDSVDKTSTLTPSTLWSSEYSEMOQLSTNSSDTESNR----- 506
XX
XX 640 MVGSKTVSQCNFYFYKRNQNLDEILQKHKEKERNARRKKKAPAAASBEAAPPV 699
XX
XX 507 -----HKLSSGL--LPKLAISTEGEQD---EASCPGDPHEEPCKPAL 544
XX
XX 700 VEDEMEASGVSGNEEMVEEAEALHASGNEV-----PRGCSGPATVNNSSDT. 748
XX
XX 545 PPEE-----CAQEEPEVTPASTISSTLSVGSFSEHLQDINGRSECVD--STDNSSKP 596
XX
XX 749 ESIPSPH-----TEAAKOTQNGPKPATIGADG-----PPGPPTPRR 788
XX
XX 597 SSEPASHMARQRLSETEKKISG-----KVTKLSASALSIMIPGDMFVSPGLSPMSPHS 652
XX
XX 789 TSRAPTEPTPASEATCAPTPAPPSPAPPPVVPKKEEEE--TAAAPPVEGE----- 841
XX
XX 653 LS-----SDPSSSRDSSPSSDASAASPHQPIVHSSGKNYGFTRAIRVYVGDSDIYT 707
XX
XX 842 -----EQKPPAAE-----ELAVDTGK-----AEEPVK 863
XX
XX 708 VHHI VNVVEGSPACQAGLKAGDLITHINGEPVHVLVHTEVIELLLKSGNKVSIITTPFE 767
XX
XX 864 SECTEABEGPAKGDAAEAATAEALKAKEKGGSGRAT-----TAKSSGAPQDSDSA 919
XX
XX 768 N---TSIKTGPARNYSYKS--RMVRSKSKSKESLERRSLRFLKAKQSPPLHTSRGF 822
XX
XX 920 TCSADEVDEAGDGKRNLLSPRLTPTGDPANASPOKPLDLKQKORAAAIPPIQVT 979
XX
XX 823 SC-----LNRSLSGSESL---PGSPHSLSPRS----- 848
XX
XX 980 KVHEPPREDAAPTKAPPAPPPQNLQPESDAPQPGSSPRGKSRSP-----APPADKEA 1034

Db 849 -----TPSYRSTPDFPS--GTNSSQSSSPSSAPNGPAGS--GHIRPSTUHLGAPKLGQR 900
Qy 1035 FAAEAQKLPDGPWCWTSGLPFPVPPREVIVKASPHA--PDPSAFSYAPPGHPLPLGLHDTA 1092
Db 901 YRSGRRKSAGNIP-----LSPLARTPSTPQPTSPQSPSPSPLLGSLG 943
Qy 1093 RVPVLP--PPTTSNPPPLISSAKHPSVLRQIGAISQMSVOLHVYPYSHAKAPVGPVTM 1150
Db 944 NSKIAQAFESKMHSPPTIVR-----HIVRPKSAEPPRSPSL-- 978
Qy 1151 GLPLPMDPKKLAPFSGVQEQQLSPRQAGPPESLGPVPTAQAEASVLRGTALGSPVGGISITK 1210
Db 979 -LKRVOSEKLSFGSGDKKHLCR-----KHSLEV-TOEE--VQREQQREAPLOSLE 1029
Qy 1211 GIPSTRVPSDSAITYRGSITHTGTPADVLYKGTITRIIG-EDSPSRILDRGREDSLPKGHVI 1269
Db 1030 NVCDVPPLSRARPEQGCL-----KRPVSRKVGQESVDDLLDR--DKL-KAKVV 1075
Qy 1270 YEGKKGHVLSYEGGMSVTOCSKEDGRSSSGPPHETAAPKRTYDMMMEGRVGRASSASIEG 1329
Db 1076 VK-----KADGF----- 1082
Qy 1330 LMGRAIPPRHSPHLKEQHHRIGSITQGIPRSYYEAOEDYLREAKLLKREGTPTPPPP 1389
Db 1083 -----PEKQESH--QKSHGPGSDLENFALKLE-----EREKKVY-----PK 1117
Qy 1390 SRDLTEAVKTOA-----LGPLKLKPAHEGLVATVKEAGRSITHEIPRELHRTPELPLA 1442
Db 1118 AVERSTFENKASMQEAPPLGSLLDKALHKQASVRASEGMSDGPVPAEHROGGGDFRA 1177
Qy 1443 PRPLKEGSITQGTPLKYDTGASTTG---SKKHDVRSILSGPGRTTFPPVHPLDMADARAL 1499
Db 1178 PAP---GTLDQGLCHSLDRIGSGKEGTEKSKQAKELLRCE-KLDSKLANIDYLRKMSL 1233
Qy 1500 ERACVYESLKSPPGTASSSGGSIARCAPVIVELGKPRQSPPLYEDHGAPFAGHLPGRSP 1559
Db 1234 EDK--EDNL-----CPVL-----KPKMTAGSHE-----CLP-GNP 1260
Qy 1560 VT---MREPTPRLOEGS-LSSSKASQDRKLTSTPREI-----AKSPHSTVPVHHPH 1606
Db 1261 VPTGCGQPPPPASESRAPVSTHAAQMSAVSFVLKALTGRVDSSTEXEPLGVAPE-SPV 1319
Qy 1607 PISPEVHLRG-----VSGVDLVRSHP--LAFDPT-----SIPR 1639
Db 1320 RKSPSEYKLEGRSVSCLEPIEGTLDIALLSQPAASKTELPSESAQSPSGDVRASVPP 1379
Qy 1640 GIPL-----DAAAYYL-PRHLAPNTYPHLYPPVLYRGYPDTAALENRQTII---NDY 1689
Db 1380 VLPSSSGKNDTTSARELSPSLKMNKSY--LLEPMFL---PPSRGLQNSPAVSLDPDEF 1434
Qy 1690 ITSQOMHNTA---TAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPV 1745
Db 1435 KDRKGPHPTARSPGTVMESNFPQREGSFKHQ---DHTTDPKLLTCLGONLHSDLAR 1490
Qy 1746 PTEGTPATAMDRLAVLPTAPQPFSSSHSSSPLSPGPGTHLTKTPTTSSSER-----ER 1798
Db 1491 P-----RCPLPEASPSREKPGLESSESGPPTARSER 1523
Qy 1799 DDDR-----ERDRDREKXIL-----TSTTVEHAPIWRPGTEQS 1834
Db 1524 SAARADTCREPSMELCFPETAKTSDNSKNLLSVGRTHPDFTYQTOAMEKA--WAPG--- 1577
Qy 1835 SGSSSGSGSGSGSGSPASHSHAHOSHPISTPTODALQORPSVLHNTGMKGIITAVEPSK 1894
Db 1578 -GKTNHKQGP--EAKPPPRDSSLHSAIGICEKELGKVR-----RGVEPKP 1621
Qy 1895 PTVLRSTSTSSVPVPAATFPFATHCPGLGTLGCVYPTLM-----EPVLLPKEAPRVARPE 1949
Db 1622 BALLARRSLQ-----PPGIESEKSEKLS--PFSLOKDGAKPE--RKEQPLQRHPS 1669
Qy 1950 R-PRADTGHAFKAPPAR-----SGLEPASSPS-----KGSEPRPLVPVSGHATIAIR 1996

Db	877	----	TPSYRTPDPFS--GTNSSQSSPSSSAPNSPAGS--GHIRFSTLHGLAPKLGGQR	928
Qy	1035	FAAEAQKLPDGPCCWTSGLFPVPPREVIKASPHA--PDPSAFSYAPGCHPLPLGLHDTA	1092	
Db	929	YRSGRRKASGNIP-----LSPLARTPSTPQPTSFQSRSPSPLLGHSLG	971	
Qy	1093	RPVLPR--PPTIGNPPPLISSAKHPSVLERIQIGAISQMSVQLHVPIYSEHAKAPVGPVTM	1150	
Db	972	NSKIAQAFPSPKMSHPPTIVR-----HIVRPKSABPRSPPL--	1006	
Qy	1151	GLPLPMDPKKLAPFSGVKGQQLSPRGOAGPSPESLGVPTAQEAASVLRGTALGSPVGGSIYK	1210	
Db	1007	-LKRVOEEKLSPSYGDKKHLCRSR-----KHSLV--TOEE--VQREQSOREAPLOSIDE	1057	
Qy	1211	GIPSTRVPDSAITRYSITHGTGPADVLVYKGTITRIIG-EDPSRLDRGREDLSLPGHVI	1269	
Db	1058	NVCDVPLSRARPVEQCCL-----KRPVSRKVGROESVDDLLDR---DKL-KAKVV	1103	
Qy	1270	YEGKKGHVLSYEGGMSVTOCSKEDGRSSGPPHETAAPKRTYDMWEGRVGRAISSASIEG	1329	
Db	1104	VK-----KADGF-----	1110	
Qy	1330	LMGRAIPPERHSPHILKEQHIRGSIOTGIPRSYVEAQEDYLBREAKLLKREGTPPPPP	1389	
Db	1111	-----PEQOESH--QKTHGPGSDLENFALPKLE-----BREKKVY-----PK	1145	
Qy	1390	SRDLTEAYKTOA-----LGPLKLKPAHGLVATVKEAGRSIHETIPRELRHTPELPLA	1442	
Db	1146	AVERSSTFENKASMOEAPPGLSLKDALHQAQSVRASEGAMSGDGPVAEHRQGGGDFRRA	1205	
Qy	1443	PRPLKEGSIQTGPLKYDTCASITTG-----SKKHdVRSLIGSGRTFFPVPHLDVMADARAL	1499	
Db	1206	PAP---GTLQDGLCHSLDRGISGKGEGTEKSSQAKELLRCE-KLDSKLANIDYLRKKMSL	1261	
Qy	1500	ERACYBESLKSRRPCTASSGGSIARGAPVTVPELGPQRQSPFLTYEDHGAPFAGHLDRGSP	1559	
Db	1262	EDK--EDNL-----CPVL-----KPKMTAGSHE-----CLP-GNP	1288	
Qy	1560	VT---MRETPRLQEGS-LSSSKASODRKLSTPREI-----AKSPHSTVPEHHPH	1606	
Db	1289	VRPTGGQOEPPPASESRAFVSSTHAAQMSAVFVPLKALTGRVDGSTEKGLVAPE-SPV	1347	
Qy	1607	PISYEYHLLRG-----VSGVDLYRSHIP--LAFDPT-----SIPR	1639	
Db	1348	RKSPSEYKLEGRSVSCLEPIETGLDIALLSGPQASKTELSPESAQSPSPSGDVRASVPP	1407	
Qy	1640	GIPL-----DAAAAYYL-PRHLAPNTYPHLYPPYLIRGYDPTAALENQTTI---NDY	1689	
Db	1408	VLPSSSGKQNDTTSARELSPSLKWNSKY--LLEPWFL--PPSRGLQNSPAVSLPDPEF	1462	
Qy	1690	ITSQOMHNTA---TAMAQADMLRGLSPRESSLALNYAAGRGIIDLSQVPHLPVLVP	1745	
Db	1463	KRDKRGHPPTARSPTWENSPQORESSPKHQ-----DHTTDPKLTLCLQNLIHSPDLAR	1518	
Qy	1746	PTPGTPTAMDRLAYLPTAQPFPSRRSHSSPSPSGGPTHLTKTPTTSSSR-----ER	1798	
Db	1519	P-----RCPLEPEASPSREKPGRESSERSGPTARSR	1551	
Qy	1799	DRDR-----ERDRDRERKSII-----TSTTVEHAPIWRPGTEQS	1834	
Db	1552	SAARADTCREPSMELCPFETAKTSDNSKLLSVGRTHRPFDPFYTOTQAMEKA--WAPG---	1605	
Qy	1835	SGSSSGSGGGSGSSRSASHAHQHSPISPRTODALQORPSVLHNTGMKIITAVEPSK	1894	
Db	1606	-GKTNHKDGGP--EARPPDRNDSLHAGIPCCEKELGVR-----RGVEPKP	1649	
Qy	1895	PTVLRSTSTSSPVRPAATFPATHCPLGGTLDGVYPTLM-----EPVLLPKPEAPRVARPE	1949	
Db	1650	EALLARSLO-----PFGIESEKSEKLSS-FPSLQDXGAKEPE--RKEQPLQRHPS	1697	
Qy	1950	R-PRADTGHAPLAKPPAR-----SGLEPASSPS---KGSEPRFLVPVPSGHATIAIR	1996	

Db	1698	SIPPPPLTAKDLSSPAARQHCSSPSHASGRECAKPTAEPPSSSPDPPKPVAAHS----	1755
Qy	1997	TPAKNLAPHASDPDPAPPASADPHREKTKQSPFSI-----QELRLSLG-----YHGSSY	2048
Db	1754	-----ESSSHKPRGPDGPPKTKHPDRSLSSQKP-SVGATKGKEPATQSLGSSSREGKGH	1808
Qy	2049	SPGG--VEPVSPVS-----SPSLTHDKGLPKHLEELDKSHLEGLRKPQPG-PVKLGGEAA	2101
Db	1809	SKSGPDVPFATPGSONKASDGIQCGEGGSPVLHTDRAPLDAKPOPTSGGRPLEVLEKPV	1868
Qy	2102	HLPHLRPLPESQSSPLLQTA PGVKGHORVVTLAQHISEVITQDVTRRHHPOQLSAPLPA	2161
Db	1869	HLPRPGHGGSEPADOKL--SAVGEK-----QTLSPKAPK-----PS	1903
Qy	2162	PLYSFFCASCPLVDLR---RPPSDLYLPPDHGAPARGSPHSGGKRSP-----PNKTSV	2214
Db	1904	TVKDCP-TLCKQTDNRQTDKSPSQ-----PAANTDRRAEGKKCTEALYAPAE GDK	1952
Qy	2215	LGCGEGIEPVSPGWTGTEPHGHSRAVYPLYRDGEQTGTEPSRMGSKSPGNTSQPPAFPSK	2274
Db	1953	LEAGLSFVHSENLKGAE RPAAGVGKGF-----EARGKGPG-PQKPP-----	1994
Qy	2275	LTESNSAMVKS KQKINCKLNTNRNEPEYNI SQPGTEIFNMPAI TGTGLMTVRSOAVQE	2334
Db	1995	-TEAD-----KFGMKRGP-----SATGQSSFRSTALPE	2022
Qy	2335	H-----ASTNMGLEAIRKALMGYDOWEESPPLSANAFNPLNASLPAAMPITAADGRS	2390
Db	2023	KSLSGSSPPETRAGVREASAASSD-----TSSAKAAGGMLELPAPSNR	2066
Qy	2391	DHILTTPGGGKK---AKVSRGRPSR-----	2412
Db	2067	DHRKAQPAGEGRTHMTKSDSLPSFRVSTLPLESHHPDPNTMGASHRDRALSVTATVGET	2126
Qy	2413	KAKSAPGLASDRPSPVSVHSEGCNRRTPLTNRWEDRPSS	2456
Db	2127	KGKDPAPA-----OPPARKQNVGRDVKPSAPN---TDRPIS	2162
RESULT 89			
ABO14747			
XX	ID	ABO14747 standard; protein; 4873 AA.	
XX	AC	ABO14747;	
XX	AC		
XX	DT	25-AUG-2003 (first entry)	
XX	DE	Novel human protein #120.	
XX	XX	Human; NOV; gene therapy; endocrine related disease; diabetes; metabolism-related disease; obesity; central nervous system disorder; Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis; schizophrenia; depression; autoimmune disorder; inflammatory disorder; psoriasis; allergy; lupus erythematosus; asthma; cancer; inflammatory bowel disease; rheumatoid arthritis; osteoarthritis; colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer; prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis; lung disease; emphysema; obstructive pulmonary disease; haemophilia; stroke; infection.	
OS	OS	Homo sapiens.	
XX	XX		
XX	XX	WO2003023002-A2.	
XX	XX	20-MAR-2003.	
XX	XX	09-SEP-2002; 2002WO-US028539.	
XX	XX	07-SEP-2001; 2001US-0318120P.	
PR	PR	07-SEP-2001; 2001US-0318130P.	
PR	PR	10-SEP-2001; 2001US-0318430P.	
PR	PR	17-SEP-2001; 2001US-0322636P.	
PR	PR	17-SEP-2001; 2001US-0322781P.	

PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 17-APR-2002; 2002US-0373212P.
PR 06-SEP-2002; 2002US-00236177.
XX (CURA-) CURAGEN CORP.
PA
XX
XX
PI Spytek Ka, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
PI Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
PI Shimketa RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
PI Riegler DK, Taupier RJ, Shenoy SG, Liu X, Padigaru M, Alsobrook JP;
PI Lepley DW, Edinger SR, Burgess CE;
XX
XX WPI; 2003-313242/30.
DR N-PSDB; ACD19440.
XX
XX
XX New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX).
PT and polynucleotides, useful in gene therapy, e.g. for treating or
PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
PT stroke or infections.
XX
XX Claim 1; Page 340-341; 586pp; English.
XX
XX The invention describes a new isolated polypeptide (NOVX). The NOVX
CC polypeptide, nucleic acid and antibody are useful as therapeutics,
CC particularly in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, which includes a pathology associated
CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
CC therapy for treating the disease or condition. In particular, the NOVX
CC polypeptide or polynucleotide is useful for treating endocrine/
CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
CC asthma, inflammatory bowel disease, rheumatoid arthritis or
CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
CC These are also useful in developing powerful assay system for functional
CC analysis of various human disorders, as well as in diagnostic
CC applications, and for monitoring the effects of drugs during clinical
CC trials. This is the amino acid sequence of a novel human NOV protein
XX
SQ Sequence 4873 AA;

Query Match 3.1%; Score 413; DB 6; Length 4873;
Best Local Similarity 19.0%; Pred. No. 1.7e-13;
Matches 477; Conservative 278; Mismatches 894; Indels 868; Gaps 105;

QY 552 DTSGDNDEKVAASKRKTANSQGRKGRITRSMANEANSEAITPQOASBLAMELNE 611
DB 44 DLSQSEERRQIAA---VMSRAOGLPKGSVPPAAA-----ESPMSHRKQLDSSHPK 94

QY 612 SSRWTEENETAKGGLLEH-----GRWSATARMVSGTKVSCQKNFVNYKK 658
DB 95 QSGRRPDPGRPAQPLGSKSRRTDTFRSEQKLPGRSPSTI-----SLKE 137

QY 659 RQNLDEILOQHLKMB-----KERNARRKKKAPAAASEAAPPVVEDEMEASGVSGNE 714
DB 138 SKSRDTLKEHKSSWMPGLSEVNA-----LSAVSSVVNKNFNDLSDSASGEETK 191

QY 715 EEWVEAEALHASG-----NEVPRGCSGPATVNNSSDTSIPSPTEAAKDTG 763
DB 192 KQKVVQEGKQPEGIKPLPQQPPKPIKQGGGRDPLQDGTGPKSISSQPEKIKS-- 249

QY 764 QNGKPPATLGADPPPGPTTPTTSRA-----PIETPT 798

DB 250 ---QPPGT---GKPIQGTOTPTDHAQLPLORDASRPQTKQADIVRGESVKSLPSP 301
QY 799 ASEATGAPTPPAPSPSAPPPVVKKEKEEETAAPVVEEGEEOKPPAAEELAVDTGKA 858
DB 302 SK-----PPIQPTPGKPPAQPFCHESQSPGAKPPAQPSGLTKP-----LAQOPGTV 349
QY 859 EEPVKSECTEAEAE---GPAK-----GKDAEAEATAEGALKAEKKEGGSGRATTAKSS 909
DB 350 KPPVQPPGTTKPPAPLGPAPKPPAQOTGSEKPSSE---OPGPKALAAQPPGVOK-TPAQOP 405
QY 910 GAOQSDSDSATCSADEVDEAEGGDKNRLSPRSLTTPGDPRANASPKPKPLDLKQLKQR 969
DB 406 G-----PAKPTQVQGTGPKPLAQOPGLQS 429
QY 970 AAAP-PIQVTKVHPPREDAAPTKAP-----PAPPPQNLOPESDAP---QOPGSS-PRG 1021
DB 430 PAKAPGPTKTPAQTKPPSQOPGSGTKPPQQOPGAKPSFQ---QPGSTKPPSQOPGSAKPSA 487
QY 1022 KSRSPAPADKEAFAAEAQKLPDPPCWTGSLPPFPVPPREVIKASPHAPDPSPASVAPPG 1081
DB 488 QOPSPAKPS-----AQPF---TKPVSGTGFKPLQPTV---SPSAKQ-----PPS 527
QY 1082 HPLP---LGLHDTARPVLPRPTIS-----NPPLISSAKHPSVLBRQ 1121
DB 528 QGLPKTICPLCWTTELLHVPKCANFNCTECQTTVCSLCGFNPNPHLTEAKEMWCLNCQ 587
QY 1122 ICALISQMSVOLHVPYSEHAKAPVQPTV-----MGLPLM-----D 1157
DB 588 MKRALGGDLAP---VPSPPQPKLTAPVTTSVAVSKSSPQOQTSPPKDAAPKQDLSKAP 645
QY 1158 PKKLAPFGVKQEQJ---SPRGOA-----GPPESLGVPTAQEAS-----1193
DB 646 PKPPPL--VKQPTLHGSPSAKAKQPPADSLSKPAPKPEPSVSEQKAPVADDKPKQ 703
QY 1194 ---VLRGTALGVSPGSIITKGIPTKVPKPS-----DSAITYRGSITHG---TPAD 1236
DB 704 KMVKPTTDLVSSSSATTKPDIPSSKVSQQAEBKTTPLKTDLSAKPSQSPPTGKVTPTD 763
QY 1237 ---VLYKCTIIRIIGEDSPSLDGRSDSLPKGHVIVYEGKHGVLVSYGSGSVTCQSKED 1293
DB 764 SKAIPRPASDSKILSHPGPSSKQKQVDP-----VOKKEEPKKAQTKMSPKDKAPM 817
QY 1294 GRSSGGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIPPRHSPHILKEQHHR- 1352
DB 818 PKGSPTP---GPRPT-----AGQTVPTPQQSP---KPEQSR 850
QY 1353 ---GSTQIGIPRSYVEAQSD-----YLRREKLLKREGTP-----1384
DB 851 FSLNLGISTDA-PKSQPTTPOETVTGKLFPGCASIFSQASNLISTAGQPGHSGSGGAP 909
QY 1385 -----PPPPSRDLTEAVKTCOALGPKLKPAAHEGLVATVKEAG 1422
DB 910 MKQAPAPSQPTSQPKSTQCAPPAKSIIPVKETKAPAAEKLEPKAE-QAPTVKRTE 968
QY 1423 RSIHEIP---REELRHTPELPLAPRLKEGSIQGT-PL-----KYDTGASTTSGK 1469
DB 969 TEKPPPIKDSKSLTAEPQKAVLPKLEKSPKPESTCLCKTELNIGSKDPNFTCTEC 1028
QY 1470 KHDVRSILIGSPGRTFPVHPLDV-----MADARALERACRYESLSKRP 1512
DB 1029 KNOVCNLGCFN---PTPHLTEIQEWLCLNCQTORAISGQLGDIRKMPPA---PSGPKASP 1082
QY 1513 GTASSSGGSIARGAPVIVPELGGKPRQSLTYEDHGAFFAGHLPKGSVPTMREPTP-----1567
DB 1083 MPVPTESSQKTAVPPQVKLVKKQEQEVKTEAEKVI-----LEKVKETLSNEKIPPMWTT 1137
QY 1568 ---RLQEGSLSSSKAS-----QDRKLTSTPREIAKS 1595
DB 1138 DQKQESKLEKDKASALQKPKPEEKLIPEEKIRSEKKKPLEEKKPTPEDKLLPE 1197
QY 1596 PHSITVPEHHPIPIPFYEHLLRGVSGVDLYRSHIPLAFDPTSIPIRGIPLDAAAAYVLPRL 1655

Db 1198 AKTSABEQKH-----DLLKSQVOIAEEK-----LEGRV 1226
QY 1656 APNPTYPHYLPYLYRGYPD--TAALLENROTIINDYITTSQOMHNTATAMAOQADMLRGL 1713
Db 1227 APKTVOEGKQPTKMEGLSGTQSLPKEDDKTKTKKEQPQPPCTAKDQEKED-----1281
QY 1714 SPRESSLALNAAAGRGIDL-----SQVPHLPVLVP-----PT 1747
Db 1282 -DKSDTSSSQPKSPQGLSDTYGSSDGLSSGLGEIPSLPTDEKILKGLKKDSFQSS 1340
QY 1748 PGTP-----ATMDRLAYLPTAQPPS-----SRHSSSLSPGCPPTH 1784
Db 1341 PSSPSOLAKLESTVLSILEAQASTLADSEKKTQPHVEPSPEQPKDQEKTSLEIT 1400
QY 1785 LTKPTTTSSSERBD--RRDRD--RRREKSGILSTTT-----VEHA 1824
Db 1401 ISSEEEKESQERKDFKDSQODIPSSKDHEKSEFVDITTRRPFYDSVESSSENS 1460
QY 1825 PIWRPTEOSSGSGS-----SGGGGG-----SSRRPASHSH-----AHQH 1860
Db 1461 PVPQRRTTSVSGSSSDEVYKQDSQSGSEEDFIRKQIIEMSADEDASGSEDDEFIRNQL 1520
QY 1861 SPISPRQDALQRPVSLHNTGKGLITAVEPSKPTVLSTSTSSVPRPAATPPATHCP 1920
Db 1521 KEISSTESQKKE-----ETKGKGITAGKHRLT--RKSTS-----1556
QY 1921 LGGTLGVPTLMEPVLLPKEAPRVARPERPRADTGHAFKAPKPARSGLEPASSPSKGS 1980
Db 1557 -----IDEDA--GRHSWHDEDEAFDESP-----ELKRYETKSQ 1590
QY 1981 PRELPVPVGHATATPAKNLAPHASPPAPPASAGDPHREKTSQKPFST-----OE 2035
Db 1591 SEELV--VTGGGLRRFKTIEL-----NSTIADKYSAESSQKTSLYFDEEPE 1636
QY 2036 LELRSL-----GYHGSYSPEGVFVPSVPSLTHDKGLPKHLELDKSH 2081
Db 1637 LEMESLTDSPEDRSRGGSSSLHASTFP---GTSPTSVSLLDED-----SDSSPSH 1685
QY 2082 LEGELPKQPGVKLGGAHLPLRPSPSS-----SPLLTQAPGVKGHRVV 2133
Db 1686 KKGESKQR-----KARHRPHGLPLPTIENSSSEELREEEELKE-----QEKREI 1733
QY 2134 TLAQHTS-----EVITQDYTRHHPOOLSAPLPAPLYSPFGACPVLDLRPP--SD 2182
Db 1734 EQQRKSSSKSKDKDELRAQRRRERPK--TP-PSNLSPIEDAS-PTEELRQAAMEE 1788
QY 2183 LYLPPDPHGAAPGSPHSSECKRSP-----PKNTSVLGGEGDIEPVSPPEGM 2231
Db 1789 LHRSSCSEYSPSIES-DPEGFEISPEKIIIEVQVKYKLPATAVSLYSPTDE--QSIMQKGS 1845
QY 2232 TEPGHSRSVYPLLYRDGEO-----TEPSRMGSKSP-----GNT 2265
Db 1846 QKALKSAEMYEMMKTHYKAFPAANERDEVEKEPLYGMLIEDYIYESLVEDTYNG 1905
QY 2266 SQPAPFSLKTESNAM-VKSKQKEINKLNTNRNEPEYINSQPGTEIFNMPPA-----2318
Db 1906 SVDGSLLTQREENGPMQKGRQKI--RLSEQIYEDPMQKITDLQKEFVELESLSVVP 1963
QY 2319 -----ITGTGLMTYSQAOVEASTNMGLEAIRKALMKGYDOWEESP 2362
Db 1964 QEDIVSSFTIIPESHEIVDLGTMTVTSEERKLLDADAAYEELMKRQOM-----QLTP 2016
QY 2363 LSANAFNPL-----NASALPAAMPITAAADGRSDHTLTSPGGGKAKVSG 2407
Db 2017 GSPQOAPTGEDMTEDMDRMPDASLTSSVLGASLTDSTSSATSLIP-----DVKITQ 2072
QY 2408 RPSRKAK-----SPAPGLASGRPPSVSVHSEGD 2439
Db 2073 HFSTEBIEEDVTDHREIQEIIAHESLILTYSEPSATSVPPSTPSLTSSVSVCTT 2132
QY 2440 NRRTPLTN-----RVVEDRPSAGSTPPFPNPLIM--RLQAGVMASTPPPP 2482
Db 2133 DSSSPITLLDSITVTYTPVDMITKPEDSEIESSSTYFPFGSIIDYPEEIASLDRTPAPD 2192

QY 2483 G-----LPAGSGPLAGPHHAWDEEPKPLCSCQVETLSDSE 2517
Db 2193 GRASADHIVISLSDMASSIIESVWPKPEGVA-----DTVSTDLLISEKDPVKKAK 2243

RESULT 90
ABR64281
ID ABR64281 standard; protein; 2468 AA.
XX ABR64281;
AC ABR64281;
XX DT 15-OCT-2003 (first entry)
XX Angiogenesis protein BNO382.

XX Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW antipsoriatic; antiarteriosclerotic; cardiac; vasotropic; angiogenesis;
KW gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; cardiovascular disease; atherosclerosis;
KW ischemic limb disease; coronary artery disease.

XX Homo sapiens.

XX WO2003027285-A1.

XX 03-APR-2003.

XX 19-SEP-2002; 2002WO-AU001282.

XX 27-SEP-2001; 2001AU-00007973.

XX 27-SEP-2001; 2001AU-00007974.

XX 11-OCT-2001; 2001AU-00008210.

XX 29-OCT-2001; 2001AU-00008532.

XX 13-NOV-2001; 2001AU-00008838.

XX 28-AUG-2002; 2002AU-00951032.

XX (BION-) BIONOMICS LTD.

XX Gamble JR, Hahn CN, Vadas MA;

XX WPI; 2003-354555/33.

XX N-PSDB; ACF34559.

XX New angiogenic genes and polypeptides, useful for diagnosing,

XX prognosticating or treating an angiogenesis-related disorder, e.g.

XX cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or

XX cardiovascular diseases.

XX Claim 15; SEQ ID NO 216; 90pp; English.

XX The invention relates to the isolation of novel genes (ACF34446-ACF34559)

XX encoding proteins (ABR64180-ABR64281) involved in the process of

XX angiogenesis. The nucleic acid molecules are useful in identifying and/or

XX obtaining full-length human genes involved in an angiogenic process. The

XX nucleic acid molecule, polypeptides or complexes encoded, cells or

XX genetically modified non-human animals derived from these are useful for

XX the screening of candidate pharmaceutical compounds used in treating,

XX angiogenesis-related disorders. They are also useful for diagnosing,

XX prognosticating or treating an angiogenesis-related disorder, which

XX involves uncontrolled or enhanced angiogenesis or is a disorder in which

XX a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,

XX diabetic retinopathy, psoriasis or cardiovascular diseases such as

XX atherosclerosis), or involves inappropriately arrested or decreased

XX angiogenesis or is a disorder in which an expanding vasculature is of

XX benefit (e.g. ischemic limb disease or coronary artery disease). The

XX modulator of expression or activity of the polypeptide encoded by the

XX nucleic acid sequence is useful for manufacturing a medicament for the

XX treatment of an angiogenesis-related disorder. This sequence corresponds

XX to one of the novel angiogenic protein

XX Sequence 2468 AA;

Db 1351 EVIEKPP-----AVP-----VSF 1363
Qy 1362 SYVEAQDYLRREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLP-----AHEG-L 1414
Db 1364 EFSDAKDN-----ERASVPMDEVPD--SESPIEKVLPLRSPPPLIGESAYESFL 1414
Qy 1415 VATVKEAGRSIHIPREELRHTELPAPRPL-----REGSITQGTPLKYDTG 1462
Db 1415 SADDKASGRGA--ESPFEKSGKQSDQVSPVSEMTSTSLYQDKQSGKSTDPAPIKEDFG 1473
Qy 1463 ASTTSGKKHUVRSLGSPG-----RTFPVHP--LDVM-----AD 1495
Db 1474 QE----KKTDDVAMSQPALALDERKLGDVSPQIDVQFSGFKEDTKMSISEGTVSDKS 1530
Qy 1496 ARALERACVEESLKSPPGTASSGGSIARGAPVIVPELCKPRQSPUTYEDHGAPFAGHLP 1555
Db 1531 ATPVDGVAEDTYSHMEGVASVSTASVATSS---FPEPTDDVSPGLHAEVSPHSSTVD 1587
Qy 1556 RGSPTVMREPTPLRLQEGSLSSKASQDRKLTGTPREI-----AKSPHSTVPEHHHPISP 1611
Db 1588 DLSLSVSVVQPTTFQETENSPSKEECPRPMISPPDFSPKTAKS--RTPVQDHRSE-----1641
Qy 1612 EHLRLGVSGVDLYRSHIPLAPDFTSIPRGIPLDAAAAYLPRHLAPNTPYPHLYPPYLIR 1671
Db 1642 -----QSSMSIEFGQESPEQSLAMDFS-----RQSPDHPTV-----1672
Qy 1672 GYPDTAALENROTILINDYITSQMHNTATAMAQRADMLRGLSPRESSLALNVAAGPRGI 1731
Db 1673 GAGVLHITENGPTEV--DYSFSDM-----QDSSLHKIPPMEEP-----SITQDN-----1715
Qy 1732 IDLSQVPHLPVLVPPPTPGTATAMDRLAYLTAPOQPFSSRHS-----SSPLSPGPGTHLPK 1787
Db 1716 -DLSEL-----ISVSQVEASPST-----SSAHTPSQIASPLQEDTLDVAP 1755
Qy 1788 PTTTS-----SBERDRDRDRDREREKSLTSTTVEHAPIWRPTEQSGSGSGSG 1842
Db 1756 PRDMSLYASLTSEKVOLEGE-----KLSPKDSISPLTPRESSPLYSPTFSDSTSAVKET 1811
Qy 1843 GGGGSSRRPASHAHQHSPISPRTODALQQPSPVLHNTGMKIITAVEPSPKPTVLRSTS 1902
Db 1812 ATCHSSSSP-----PIDAASAEYGFASVLFDTMQHHLALNRLDTPGLEKOSG 1861
Qy 1903 TSPVRAPAATFPFATHCPGLGTLGVYPTLMBPVLLPKEA-----1942
Db 1862 GKTP-----GDFSAYQKPEETTRSPDEEDYDSEYKTTTSDVGGY 1905
Qy 1943 -PRVARPERPRADTGAF--LAKPPARSGLEPASPSKGEPRPLVPVPSGHATTARTPA 1999
Db 1906 YEKIERTTKSPSDSGSYETIGK-----TTKTPEDGDYSYETIEKTT-----RTPE 1951
Qy 2000 KNLAHPHASPDPAPPASADPHREKTQSKPFSIQLELRLSLGYHSGSYSPGVEPVSFV 2059
Db 1952 EGGYSYDISEKTTSP--EVSGYSYKETERSRLLDDI-----SNGYDDSE---DG-----1997
Qy 2060 SSPSLTHDKGLPKHLBELDKSHLEGELRPKQCPVVKLGCEAAHLPHLRPLPSPQSSSPL 2119
Db 1998 -----GHTLGDPSYSYETIEK-----ITSPSESGVS---2024
Qy 2120 LQAPGVKGHQVRVUUAQHISEVITQDYYRHHPQQLSAPLAPLPSFPGASCPVLDLRP 2179
Db 2025 YETSTKTRTPDTSTYCYETAETKITR-----PQ-----ASTSYETSJDLCTYAEKKS 2072
Qy 2180 PSDL-----VLPDPDHGAPARGSPHSEGGKSPSPNKTSVLGGEDGIEPVS 2226
Db 2073 PBEARQDVLCVSSCEYKHKPTLSFSPFINGNPLEWFASBPTESEKPLTQSGAP---2130
Qy 2227 PPEGTEPHGHSRAVYLLRYRGEQTEPRSRMGSKSPGNTSQ--PPAFFSKLTESNMAVK 2284
Db 2131 PPPGGKQOG-----RQCDTEPTVSSESAPSDQSDVPE-----TE-----2167
Qy 2285 SKQKINKKLNTHNRNEPNYSIQPCTEINFNPATITGTLMTYRSO-----AVQE-----2334
Db 2168 -ECPSITADANIDSESE-----TIPTDKTVTKHMDPPPPAPVQDRSPSP 2212

Qy 2335 -HASTNM-----GLEAIIRKALMGKYDQWESPPLSANAFNPLNASASLPAAMPITAAD 2387
Db 2213 RHPDVSMVDPEALATEQNLKAL--KKDLKEKT-----KTKKPGTKTKSSSPVKKSD 2262
Qy 2388 GRSDHILTSPPGGGKAKVS---GRPSSRKAKSPAPGLASGDRPPSVSVSHSGDCNRRTP 2444
Db 2263 GSKSPLAASPAGLKESDVKSRVASPKKESVEKAAKPTTTTPEVKAARGEKCKETKN 2322
Qy 2445 LTNRWEDRPSAGSTPPPNPLIMRLQAGVWASPPPCLPAGSGPLAGPHEA 2497
Db 2323 AANASASKSAKTATAGP-----GTTKTKSSAVPPGLPVYLDLCYIPNHS 2367

RESULT 94

ADBE62715
ID ADE62715 standard; protein; 2468 AA.
XX AC ADE62715;
XX AC AC
DT 29-JAN-2004 (first entry)
XX Human Protein NP_005900, SEQ ID NO 8648.
DE Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; NP_005900.

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating

Db 2264 RHPDVSMDPEALAEQNLGKAL--KKDLKEKT-----KTKKPGTKTKSSSPVKKSD 2313
Qy 2388 GRSDHTLTSPGGGKAVS--GRPSRKAKSPAGGLASGDRPSPVSVHSEGDCHRRTP 2444
Db 2314 GKSPPLAASPAGLKESSDKSVASFPKKESVEKAAKPTTTPVKAARGESKOKETKN 2373
Qy 2445 LTRNWBDRPSSAGSTFPYNYNPLIMRLQAGVMASPPPPGLPAGSGPLAGPHHA 2497
Db 2374 AANASAKSATAAGP-----GTTKTKTSAYPPGLPVYLDLCYIPNHS 2418

RESULT 96
ID ABO44398
AC ABO44398 standard; protein; 2185 AA.
XX ABO44398;
XX
DT 26-SEP-2003 (first entry)
XX Novel human protein kinase #18.
DE Humán; kinase; enzyme; cosmetic application; nutraceutical application.
KW Homo sapiens.
XX
OS
XX
PN US6541252-B1.
XX
XX 01-APR-2003.
XX 14-MAY-2001; 2001US-00854856.
XX 19-MAY-2000; 2000US-0206015P.
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Walke DW, Hilbun E, Doncho G, Turner CA;
XX WPI; 2003-575927/54.
DR N-PSDB; ACH03781.
XX
XX New nucleic acid encoding novel human proteins, useful in cosmetic and
PT nutraceutical applications.
XX
PS Disclosure; Page; 11pp; English.
XX
XX The invention relates to a new isolated nucleic acid encoding a novel
CC human protein kinase. The nucleic acid is useful in cosmetic and
CC nutraceutical applications. The present sequence represents the amino
CC acid sequence of a novel human protein kinase. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6541252B1
XX
SQ Sequence 2185 AA;

Query Match 3.1%; Score 408; DB 7; Length 2185;
Best Local Similarity 19.6%; Pred. No. 1.1e-13;
Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;
Qy 147 PVSPPPHHTDELELVPLSKELIQNDVRDREITMVEQOISKLKKQOOLSEBAK 206
Db 39 PLSLQPSIPAAPQSPAPPEHRETV-----TATTSQVAQPPAAAPGEQAVA 89
Qy 207 PPEP-----EKVSPPPPIESKHSILVQIYDENRKAEEAHHLEGLQPQVELPLY 257
Db 90 GPAPSTVPSSTSKDRPVSPSL-----VGSKEEPPA 121
Qy 258 NQPSDTRQYHENIKINQAKKLLILFKRNHARKQWKQFCQVQDQMLEAKKVERIE 317
Db 122 RSGSG-----GGSAPKQPEERSQOQDDI-ELETRAVGMS 155
Qy 318 NNPRRRAKSKV-REYERQKPFPIRQR-----ELQERMQRVQQRGSGLSMAARSE 369

Db 156 NDGRFLPKDIEIGRGSFKTVYKGLDTETTVAVWELQDRK-----LTKSERQRP 205
Qy 370 HEVSEIIDGLSQENLEKOMROLAVIPMLYDADQORIK-----FINMGLM-ADPMKVYK 424
Db 206 KEEAEMLKGL-QHPNIVR-----FYDSWESTVTKKKCIVLVTLMSTGLTKYL 253
Qy 425 DR-----QVMWMSOEKETREKFMQHPK-----NFGLIA 455
Db 254 KRPVWKIKVLRWCROILKGLQFLHTRTPPIIHRDLKCDNIFITGPTGSKVIGLGL-- 311
Qy 456 SFLEKRTVAECVL-----YYLTKKNENY-----KSLVRRSYRRRGSKSQOQOQOQ 501
Db 312 ATLKRASPAKSVIGTPEFMAPEWMEYBEKYDESVDVAFGCMCLEMATSYFPYSECQNAQI 371
Qy 502 QOQOQOQOQOQ-----MPRSSQ-----EKDEKEKEKEABEKEEPEVENDKEDLLK 547
Db 372 YRRVTSGVKPASFDKVAIPEVKEIIEGCIQONKDERYSIKOLLNHAFFQETGVRVELAE 431
Qy 548 EKTDDTSGEDNDEKAVASKGRKKTANSQGRKGRITRSMANEANSE-EAITPOQSAELAS 606
Db 432 E-----DDGEKIAIKLWLRIEDIKKLGKYKDNEAIEFSPDLERDVPEDVAQ-- 478
Qy 607 MELNESSRWTEEMETAKKGLIEHGRNWSAIRMVGSKTVSQCKNFYFNKRONLDBIL 666
Db 479 -EMVESGYVCEGDHKTMAIKDR-----VSLIK-----RKGEORQLVR 516
Qy 667 QOKLKMEKERNARKKKA-----PAAASEAAFPVVEDEMEASG 709
Db 517 BEQEKKEEESLQOQVEQSASQOTGIKQLPSASTGIPASTSTASVSQVPEPEPEA-- 574
Qy 710 VSGNEEMVEEAALHASGNEVPRGECGSPATVN--NSSDTESIPSPHTEAKOTQNGP 767
Db 575 -DHOQLOVQOPSISVLSGDIVDSGQGSVFTESRVSSQQTIVSYGSH-EOAHTGTVP 632
Qy 768 KPPATLGADGPPPG--PPT-----PPR 787
Db 633 HIPSTVQAQSPHGVYPPSSVAQSQSQSSSLTGVSSQPIQHPOQOQIQOTAPFPQ 692
Qy 788 RTSRAPIEPT-PASEATCAPTPPAPPSPAPPVPVPEKEEKEEETAAAPPVE--EGEQK 844
Db 693 QTVQVLSQTSSTSEATTA-----QPVSQOAPQVLPQVSAQKQLPVSQVPTIQEPQI 747
Qy 845 PPAAEE--LAVDTGKAEPVKSECTEEAEFGPAKDAEAAEATAEGALKAKEKGGSGR 902
Db 748 PVAQPSVVPVHSGAHLFV-----GQPL 771
Qy 903 ATTAKSGAPQDSDSSATCSADEVDEAGGDKNLLSPRSLLTPTGDPANASPOKPLD 962
Db 772 PTPL-----LPQYVPSQIPSTPHVSTAQTG-----FSSLPTMA-----AGITQPLL 814
Qy 963 LKQLKQRAAATP-----PIQVTKVHEPPREDAAPTKPAAPPNPNLOPESAPQOPG 1016
Db 815 TLASATTAALPGVSTVVPSQLPTLLQFVTO-----LPSVHPQLLQ--AVOSMG 863
Qy 1017 SSPRGKSRSPAPPKAEFAAEAKLPCDPPCWTSGLPFPVPPREVIKASPHADPSAFS 1076
Db 864 I-----PANL-QAAAEVPLSSGD--VLXQGFPPRLPPQY-----PGDSN 899
Qy 1077 YAPGCHPLPLGLHDTAPVL--PRPPTISNPPPLISSAKHPV--LERQIGATISQGMV 1131
Db 900 IAPSSNVASVCIHST--VLXPPMPTVLATPGVFTVVOYVBSNLLVPMGGV--GGQV 954
Qy 1132 QLHVYPYSEHAKAPVGTWGLPMDPKLAPFSGVQKQELSPPGAGPPESLGV-----1186
Db 955 QVSQFGSLAQAPTSSQAV-----LESTGV-----SQVAPAEVAVAOQA 998
Qy 1187 --PTAQEASVLRGALSGVPGSITKIGPSTRVPSDSAITVRGSIHTGTPADVLYKGTIT 1244
Db 999 TOPTTLASSV--DSAHSDVASG-MSDG--NENVPSSSG-----RHEGRTT 1038
Qy 1245 RIIGEDSPSLDRGREDSLPKGHVYISGKGHVLSYEGGMSVTCQSK-----1292

Db 1039 KRYHRSVRSRSHKTSRPLRLNLSNKG-----RVVECOLETHNRKMTVTKFD 1090
QY 1293 -DGRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPPPERHSPHLKEQHII 1351
Db 1091 LOGDN-----PEETATWVNDPI-----LAIERESFVQVREII-----EKADEMLSEDSV 1138
QY 1352 RGSITGIPRSYVEAGEDYLRREAKLLKREGTPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1411
Db 1139 EPEGDQGL--ESLQGHDDYGFSGSKLEGEFKQIPASSM-----PQIGIPT 1184
QY 1412 EGIIVATVKAERS--IHEIPRELRITPELPLAPRLKESITGQTPKLYVDYDCASTTGSK 1469
Db 1185 SSLTQVHSGRRFVSPVESLRSKVPP-----SEIT-----DTVAASTAOS 1229
QY 1470 -----KHDVRSL-----IGSPGRTPFPVPHPLDVNMADARALE 1500
Db 1230 PGNLSSHASSLSLQAFSELRAQTEGNTAPPNFSHTGTPFPVPP----- 1278
QY 1501 RACYESLSKSPGTASSSGSIARGAP-----VIVPELGKPRQSL--TYEDHGAPPA 1551
Db 1279 ---FLSSIAGVPTTAATAAPVATSPPNNDISTSVIQSEVTVPTERGIAGVATSTGVVTS 1335
QY 1552 GHLPRGSPVMTREPTRLQEGSLSSSKAQDRKLTSTPREIA---KSPHSTVPEHHHPPI 1608
Db 1336 GGL-----PIPPVSESPLVSSVS-----ITIPAVVISITSPSLQVPTSTSEIV 1381
QY 1609 SPVEHLLRGVSGVDLYRSHRPIAFDPTSIPIRGIPDLDAAYLPHLAPNPTVPHLYPPV 1668
Db 1382 -----VSSALYPS-----VTVSATASAGGSTATPCPK-----PPA 1413
QY 1669 LI-----RGYPDTAALENRQ-----TIINDYITSQQMH 1696
Db 1414 VVSQQAAGSTTVGATLTSVSTTTSFSTASQLSLSSTSTPTLAETVVSASHLSDKTS 1473
QY 1697 HNTATAMA-----QRADMLRGLSPR-----ESSIALNYAAGPRGID 1733
Db 1474 HSTTGLAFSLASAPSSSSPGAGVSYI SQPGHPLVIPSIVTASPIILPQAGPTSTPL 1533
QY 1734 LSQVPHPLVLPPTGTPATAMDRLAYLPTAPOPFSSRHSPLSPGCP--THLTKTPTTS 1792
Db 1534 LPQVPSIPLVQVANPAV--QOTLIHSQOP-----ALLPNQPHTHCP----- 1576
QY 1793 SSERDRDRDRDREREKSILTSITTVHEAPIWRPGTEQSGSGSGSGSGSSSRPA 1852
Db 1577 -----EVDSDTOPKAPGIDDIKTLEE-----KLRSLSFSEHSSGA----- 1611
QY 1853 SHSHAHQSPISPTODALQORPSVLHNTCMKI--ITAVEPSKPTVLRSTSSSPVRPA 1911
Db 1612 -----OHASVSLTS-----LVTESTVTGIPITTAVAPSK--LLTSTT-----S 1648
QY 1912 TFPPTHCPGLGGLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEP 1971
Db 1649 TCLPPTNLPLG-----TVALPVTVPVTPGQVSTPVTSTTSVGPV 1687
QY 1972 ASGSPSGSPRLPVVPSVGHATARTPAKNLAPHASPPPPAPASAPHREKTSQ-KP 2030
Db 1688 GTAPSKPPLTKAPVLPVGTLPAGTLPSQOL-----PPFPGPSL--TQSQOP 1732
QY 2031 FSTQELERSLGLVHGSSYSPEGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQ 2090
Db 1733 LEDLDAQLR-----RTLSPXITVTSV----- 1755
QY 2091 PGVPLKGGEAHPLHPLRIPE-----SPPSSSPLLQATPV-----KGHVVVTLAHI SEVI 2143
Db 1756 -GPVSNAAFTA--ITEAGTQDKGVSVQKEGCVLATSSGAGVFMGPFQVSVAA----- 1806
QY 2144 TDYTRHHPQOLSAPLAPLYSPGASCPVLDLRRPPSDLYLPPP-----DHGAP 2193
Db 1807 --DGAQKEGKNKSEDAKSVHFESSTESSVLSSSPSESTLVKPEPNGITIPGISSDVPS 1864
QY 2194 ARGSPHSEGGKRSPEPNKTS-----VLGGEDGIE-----PVSPFPGMTE 2233
Db 1865 AHKTTASEAKSDTGPTKVGRFOVTTTANKVGRFSVSKTEDKITDTKKEGPPVPPFMDL 1924

QY 2234 PGHSRAVYLLYRDCEQTEPGRMGSKSPGNTSQP--PAFFSKLTESNAAVWSKKQEI NK 2292
Db 1925 EQAVLPAPVPIKKEKP--ELSEPSHLN---GPSDDPEAAFLSRDVEDDGGSGSPHSPHQLSSK 1979
QY 2293 KLNTNRNPEPYNISQPGTEIFNMPAITGTGLMTPRSQAVQ-----EHASTNMGLEA 2344
Db 1980 SL-----PSQNLQSLSNSFNSSYMSSDNESDIEDDLKLELRKLRDKHLKEIQDLOS 2032
QY 2345 IIRKALMGKYDOWESPPPLSANAFNPLNASASLPAAMPITAAADGRSDHTLTSPGGGKAK 2404
Db 2033 ROKHEIESLYTKLGKVP-----AVIIPPAAPLS---GRRRRTKS-----KGS 2073
QY 2405 VSGRPSRSKAKSP-APGLASGDRPPSV-----SVHSEGD 2438
Db 2074 KSSRSSSLGNKSPQLSGNLGSGAASVLHPQOTLHPGPN 2112

RESULT 97

ABO44382
ID ABO44382 standard; protein; 2245 AA.

XX ABO44382;

XX AC ABO44382;

XX DT 26-SEP-2003 (first entry)

XX XX Novel human protein kinase #2.

XX DE Human; kinase; enzyme; cosmetic application; nutraceutical application.
XX KW Human; kinase; enzyme; cosmetic application; nutraceutical application.
XX OS Homo sapiens.

XX PN US6541252-B1.

XX XX 01-APR-2003.

XX PF 14-MAY-2001; 2001US-00854856.

XX PR 19-MAY-2000; 2000US-0206015P.

XX XX (LEXI-) LEXICON GENETICS INC.

XX PI Walke DW, Hilbun E, Donoho G, Turner CA;

XX DR WPI; 2003-575927/54.

XX DR N-PSDB; ACH03765.

XX XX New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.

XX PS Disclosure; Page; 11pp; English.

XX CC The invention relates to a new isolated nucleic acid encoding a novel human protein kinase. The nucleic acid is useful in cosmetic and nutraceutical applications. The present sequence represents the amino acid sequence of a novel human protein kinase. Note: The sequence data for this patent did not from part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6541252B1

XX SQ Sequence 2245 AA;

Query Match 3.1%; Score 408; DB 7; Length 2245;

Best Local Similarity 19.6%; Pred. No. 1.1e-13;

Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;

QY 147 PVSPSPPTHTDELELVPPRLSKBELIQNMDRVREITVWEQOISKLKKKQOOLEBEAK 206

Db 99 PLSLPQPSIPAAVPQSAPEPHRETV-----TATATSOVAQPPAAAAAAGEQAVA 149

QY 207 PREP-----EKPVSPPIESKRSILVQIYIDENRKAAEAHRLSLGPGVLEPLY 257

Db 150 GPAPSTVPSSTSKDRVPSQPSL-----VGSKEPPPA 181

258 NPSDTRQYHENIKINQAMRKLILYFKRNHARKQWKQKFCORYDQALMEALEKKVERIE 317
 182 RSGSG-----GGSAKEPQBERSQOQDDI--ELETKA VMS 215
 318 NNPRRRAKESKV--REYKEKOPPEIRKOR-----ELQERMQSRVQORSGGLSMSAARSE 369
 216 NDCRFLKFDIEIGRSGFKTYKGLDTETTVEVAWCELQDRK-----LTKSRQR 265
 370 HEYSEIIDGLSEGENLEKQMLAVIPPMYLDADQORIK-----FINNGLM-ADPMKYK 424
 266 KEAEMLKGL-QHPNVR-----FYDSWESTVKGKCCIVLVTMLTSGTLKTYL 313
 425 DR-----QVMNWSQEKETFREKFMQHPK-----NFGLIA 455
 314 KRFKWKIKVLRSCWQQLKGLQFLHTRTPPIIHRDLKCDNFIITGPTGSVKIGDLGL-- 371
 456 SFLEKTVACVL-----YYLTKQENY-----KSLVRSYRRRGKSGOQOQOQO 501
 372 ATLKRASFASKSVIGTPEFMAPEYEBKYDESVDVYAFGCMLEMATSEYPYSECQNAQI 431
 502 QOQOQOQOQOQ-----MPRSSQ-----EEKDEKEKEKEKEKEKEKEKEKEDLLK 547
 432 YRVTSQVKSPDKVAIPEVKIIEGCIRQNKDERYSIKDLNLHAFQOETGVRVELAE 491
 548 EKTDDTSGEDNDEKEAVASKRKTANSQCRGRITRSMANEANSE--EAITPQOQSAELAS 606
 492 E-----DDGEKIAIKMLRIEDIKLKGKYKDNEAIEPSPDLERDVPEDVAQ--- 538
 607 MELNESSRWTEEMETAKGLLEHGRNWSAIAWVGSKTVSQCKNPFYFNKRONLDEIL 666
 539 -EMVESGYCEGHKTKMAKAIDR-----VSLIK-----KRQORQLVR 576
 667 QQHKLMXEMERNARRKKKA-----PAAAGEEAAFPVPPVDEEMEASG 709
 577 EEQEKKKQESSLKQVQESSASQGIKQLPSASTGIPTASTTSASVSTQVPEPEEA-- 634
 710 VSGNEEMVEEABALHAGNEVPRGECGPATVN--NSSDTSIESPPHTEAAKDTQNGP 767
 635 -DOHQLOLQOQPSISVLSGTVDSGQSGSVFTESRVSSQQTVSYSQOH--EQAHSHTGTV 692
 768 KPPATIGADGPPPG--PPI-----PRAAAGEEAAFPVPPVDEEMEASG 787
 693 HPSVTVQASQPHGVTVPSVSVAGQSQGQOQFSSSLTGVSSQPIOHQOQOQGIQTAPQ 752
 788 RTSRAPIET--PASEATGAPTPPPAPPSAPSAPPVVPVPEKEEKEEETAAAPPVE--EGEOK 844
 753 QTVQYSLSTQTSSEATTA-----QVVSQPAQVLPQVSAKQLPVSQPVFTIQGEPI 807
 845 PPAABE--LAVDTGKAEPPVKSCTEAEAGPAKGDABAEATAEGALKAEKKEGGGR 902
 808 PVATQPSVVPVHSGAHFLV-----GQPL 831
 903 ATTAKSSGAPQSDSATCSADEVDEAGDKNRLSPRSLTTPGDDPRANASPKPLD 962
 832 PTPPL-----LPQVPSQIPISTHVSTAQNG-----FSSLPIWA-----AGITQPLL 874
 963 LKOLKORAAIP-----PIQVTKVHPREDAPATKPAAPPQPNLQPSDAPQOFG 1016
 875 TLASSATTAAIPGVSTVPSQLPTLLQPVTO-----LPSQVHPQLLP---AVQSMG 923
 1017 SSPRGKSRPAPADKEAFAEAQKLPDPPCWTSGLPPFPVPPREVIVKASPHAPDPSAFS 1076
 924 I-----PANL-GQAAEVPVLSGGD--VLYQGFPPRLPPQY-----PGDSN 959
 1077 YAPPGHPLPLGHDTPARVPL--PRPTISNPPPLISSAKHPSV---LERQIGAIQSGMSV 1131
 960 IAPSSNVASVCIHST---VLXPPMTEVLATGYFTVVQPVVSNLLVPMGV--GGV 1014
 1132 QHVYSEHAKAPGVPTVMGLPLPMDPKLAPFGVQKQEQSLSPROAQGPPELSGV----- 1186
 1015 QVSPGGSQAQAPPTSSQOAV-----LESTQGV-----SQVAPAEVAVAQQA 1058

1187 --PTAQEASVLRGTALGSPVGGSIKTIPTSRVPSDSAITYRGSITHGTPADVLYKGTIT 1244
 1059 TOPTTILASV--DSASHSDVASG-MSDG--NENVPSSG-----RHEGRIT 1098
 1245 RIIGEDSPRLDRGREDSLPKGHVITYEGKKHVLSEYEGGMSVTQCSKE----- 1292
 1099 KRHYRKSVRSRSRHEKTSRKLRIILNVSNKGD-----RVVEQOLETHNRKMTVFKFD 1150
 1293 -DGRSSGPPHETAAPKTYDMWMEGRVGRAISSASIEGLMGRAIPPEHRSPHLKEQHHI 1351
 1151 LDGDN-----PEIATIMVNDFI-----LAIERESFVQVREII--EKADEMLSEDSV 1198
 1352 RGSITQGISRVSVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTOALQPLKLDKAAH 1411
 1199 EPEGQOGL--ESLQKDDYGFSGSKLGEKQPIPASSM-----PQOIGPT 1244
 1412 EGLVATVKEAGRS--IHEIPREELRHTELPLAPRLKEGSIOTGTPLKYDTGASTTSGK 1469
 1245 SSLTQVHVSAGRRFTVSPVPSRLRESKVPF-----SEIT-----DTVAASTAQS 1289
 1470 ----KHDVRSI-----IGSPGTEPPPHPLDVMADARALE 1500
 1290 PGWNLHSHASSLSLQOAFSELRRAQMTGPNTPNFSHTGTPFPVPP----- 1338
 1501 RACYEESLKSRTGTAASSSGGIARGAP-----VIVPELKGPROSPL--TYEDHGAPFA 1551
 1339 ---FLSSIAGVPTTAAATAPVATSPNDISTSVIQSEVTVPTGEGIAGVATSGVVS 1395
 1552 GHLPGRSPVWREPTPRLOEGLSSSSKASQDRKLTSTPREIA---KSPHSTVPEHHPHI 1608
 1396 GGL-----PIPPVSESPVLSSVSS---ITIPAVVSISTTSPSLQVPTSTSI 1441
 1609 SPYEHLLRGVGDVLYRSHIPLAFDPTSI PRGIPDLAAAAYVLPRLAPNPTPHLYPY 1668
 1442 ----VSTALYPS-----VTVSATSASAGSTATPGK-----PPA 1473
 1669 LI-----RGYDPTAALENRQ-----TIINDYITSQMH 1696
 1474 VVSQQAAGSTTVGATLTSVSTTTSPSTASQLSLSSTSTPTLAEVTVVSAHSLDKTS 1533
 1697 HNTATAMA-----QRADMLRGLSPR-----PSSLALNTVAAGPRGIID 1733
 1534 HSTTGLAFSLAPSSSSSGAGVSSYISQPGGLHPLVIPSVIASTPILPOAAGTSTPL 1593
 1734 LSQVPHLPVLPPTGTPATAMDLAYLPTAQPFSSRSHSSPLSPGQP--THLTKPTTS 1792
 1594 LPQVPSIPLVQPVANVPV--QOTLIHSQOP-----ALLPNQPHTHCP----- 1636
 1793 SSERDRDRDRDREREKSIILSTTTTVEHAPIWRPGTEQSSGSSGSSGSSSRPA 1852
 1637 -----EVDSDTOPKAPGIDDIKTLEB-----KLRSLFSEHSSGA----- 1671
 1853 SHSHAHQSPISPTQDALQORPSVLHNTGKGI--ITAVEPSKPTVLRSTSSPVPA 1911
 1672 -----QHASVLETS-----LVISTVTPGIPPTTAVAPSK--LTSIT-----S 1708
 1912 TFPATHCPGLGGLDGVYPTLMPEVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEP 1971
 1709 TCLPPTNLPLG-----TVALPVTVPVTPGVSTPTVSTTSGVKP 1747
 1972 ASSPSKSEPPPLVPPVSGHATIAARTPAKNLAPHASPDPPAPPASADPHREKTS--KP 2030
 1748 GTAPSKPLTKAPVLPVGTLPAGTLPSEQL-----PPFPFSL-----TQSQP 1792
 2031 FSIQELERLSLGYSSSVSPGEPVSPVSPSLTHDKGLPKHLELDKSHLEGELRPK 2090
 1793 LEDLDAQLR-----RTLSPEXITVTSV----- 1815
 2091 PGPVKLGGEAHLPHRLPPE---SQPSSGPLLOTAPGV---KGHORVVTLAQHISEVI 2143
 1816 -GPVMAAPTA--ITEAGTQPKGVSQVKEGVLATSGAGVFKMGFRQVSVAA----- 1866
 2144 TQDTRHHPQOLSAPLAPLYSFPGASCPLVLDLRRPSPDLVLP-----DHGAP 2193

Db 1867 --DGAQKEGKNSDEAKSVHFESSTSESSVLSSTSSPESTLVKPEPNGITIPGISSDVPS 1924
 QY 2194 ARGSPHSEGGKSPENKTS-----VLGGEDGTE-----PVSPPEGMT 2233
 Db 1925 AIKTTASEAKSDTGQTKVGRFOVTTANKVGRFSVKTEKIDTKTKGVPVSPFMDL 1984
 QY 2234 PGHSRAVYLLYRDGEOTEPSPMGSKSPGNTSQP--PAFFSKLTESNSAMVSKKQEIWK 2292
 Db 1985 EQAVLPVAVIPKKEKP--ELSEPSHLN---GPSSDPEAAFLSRDVGSGSPHSPHQLSSK 2039
 QY 2293 KLNTHRNPEPEYNISQPGTEIFNMPAITGTGLMYRSQAVQ-----EHASTNMGLEA 2344
 Db 2040 SL-----PSQNLSSQLSNSFNSMSSDNESDIEDKLELRRLRDKHLKEIQDLQS 2092
 QY 2345 IIRKALMGKYDOWEESPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGGGKAK 2404
 Db 2093 RQKHETESLYTKLGKVP-----AVIIPPAAPLS-----GRRRRPTKS-----KGS 2133
 QY 2405 VSGRPSSRKAKSP-AFGLASGDRPPSV-----SSVHSEGD 2438
 Db 2134 KSSRSSSLGNKSPQLSGNLGQSAASVLPHPQOQLHPPGN 2172

RESULT 98

ABO44397
 ID ABO44397 standard; protein; 2322 AA.

XX ABO44397;

DT 26-SEP-2003 (first entry)

DE Novel human protein kinase #17.

XX Human; kinase; enzyme; cosmetic application; nutraceutical application.

OS Homo sapiens.

XX US6541252-B1.

XX 01-APR-2003.

PF 14-MAY-2001; 2001US-00854856.

PR 19-MAY-2000; 2000US-0206015P.

XX (LEXI-) LEXICON GENETICS INC.

PI Walke DW, Hilbun E, Donoho G, Turner CA;

DR WPI; 2003-575927/54.

DR N-PSDB; ACH03780.

XX New nucleic acid encoding novel human proteins, useful in cosmetic and

PT nutraceutical applications.

XX Disclosure; Page; 11pp; English.

CC The invention relates to a new isolated nucleic acid encoding a novel
 CC human protein kinase. The nucleic acid is useful in cosmetic and
 CC nutraceutical applications. The present sequence represents the amino
 CC acid sequence of a novel human protein kinase. Note: The sequence data
 CC for this patent did not from part of the printed specification but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocId=654125281

XX Sequence 2322 AA;

Query Match 3.1%; Score 408; DB 7; Length 2322;

Best Local Similarity 19.6%; Pred.No. 1.2e-13;

Matches 513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;

QY 147 PVSPSPPTDPELELVPPRLSKEELIQNMDRVDREITWVEQQISKLKKKQOQLEBAAK 206

Db 39 PLSLPQPSIPAAVQSPAPPEPHEETV-----TATATQVAAQPPAAAPGQAVA 89
 QY 207 PREP-----EKVSPPIESKHSRLVQIYIDENRKKAAAHRLGLEGLQVQVELPLY 257
 Db 90 GPAPSTVPSSTSKDRPVSPSL-----VGSKEPPPPA 121
 QY 258 NOPSDTROYHENIKINQAMRKKLILYFKRNRHARKQWKQFCORYDQLEALEKKYERIE 317
 Db 122 RSGSG-----GGSKEPQERSQOQDDI--BELETKAVGMS 155
 QY 318 NNPRRAKESKV--REYKQFPEIRKQ-----ELQERMOSRVQORSGLSMSARSE 369
 Db 156 NDGRFLKFDIEIRGSRFKTVYKGLDTETTVVAVMCELQDRK-----LTKSRORF 205
 QY 370 HVSSEITIDGLSQENLEKQMRQVLAVIPMLYDADQORIK-----FINMNGML-ADPMKVYK 424
 Db 206 KBAEMLKGL-QHPNIVR-----FYDSWESTVAGKKCIVLVTELMTSGTUKTYL 253
 QY 425 DR-----QVMNMWSEQEKETFREKFMQHPK-----NFGLIA 455
 Db 254 KFKVMKIKVLSRWCQRILKGLQLHLTRTPPIIHRDLKCDNIFITGPTGSKVIGDGLG-- 311
 QY 456 SFLEKRTVAECVL-----YYLTQKQENY-----KSLVRRSYRRRGKSKOQOQOQ 501
 Db 312 ATLKGASFAKSVIGTPEFMAPEMYBEKYDESDVDVYAFGCMLEMATSEYPSYSECQNAQI 371
 QY 502 QOQOQOQOQOQ-----MPSRSQ-----BEKDEKEKEKEKEKEKEKEKEKEKEDELLK 547
 Db 372 YRRVTSGVPASFDKVAIPEVKEIIEGCTRONKDERYSIKDLLNHAFFQETGVRVELAE 431
 QY 548 EKTDDTSGEDNDEKAVASKGRKTANSQGRKRGRITRSMANEANSE-EAITPQOQSAELAS 606
 Db 432 E-----DDGEKIAIKLWRIEDIKKLGKYKDNEAIEFSPDLERDVPDVAQ--- 478
 QY 607 MELNSSRWTEEMETAKGGLLEHGRNWSAIAARMVGSKTVSQCKNFYFNKKQKQNDL 666
 Db 479 -EMVSGYVCEGDHKTMAKAIKDR-----VSLIK-----RKREORQLVR 516
 QY 667 QQHLKMEKERNARRKKKA-----PAAASEEAAFPVVEDEMEASG 709
 Db 517 EEQEKKKQESSLKQOQVESSASQGIKQLPASGTGIPASTTSASVSTQVEPEPEA-- 574
 QY 710 VSGNEEMVVEAEALHASGNVPRGECSPATVN--NSSDTESIPSPHTEAAKDTQNGP 767
 Db 575 -DQHQQLQYQPSISVLSGDTVDGSGSSVFTESRVSSQTVSYGSGH-EQAHSTGTVP 632
 QY 768 KPATLGADGPPPG--PPT-----PPR 787
 Db 633 HIPSTVQAQSPHGVYPPSSVAQSGOSQOQSPSSSLTGVSSSQPIQHPOQOQGIQTAPPQ 692
 QY 788 RTSRAPIPT-PASEATGAPTPPPAPPSAPPVVPVPEKEEETAAAPVE--EGEOK 844
 Db 693 QTVQVSLSTSTSSRATTA-----QVSOQAPQVLPQVSAQKQLPVSPVPTIQEPEQI 747
 QY 845 PPAABE--LAVDTGKAEPEVKSECTEAEPEGPAKGDABAEABATARGALKAEKKGSGR 902
 Db 748 PVATQPSVVPVHGAHFLPV-----GQPL 771
 QY 903 ATTAKSSGAPQSDSSATCSADEVEAEGDKNRLSPRLTPTGDPNANASPKPLD 962
 Db 772 PTPL-----LPQVPVSIPISTPHVSTAQTG-----PSSLFITMA-----AGITQPLL 814
 QY 963 LKQLKQRAAIP-----PIQVTKVHEPREDAAPTAPPPPPQNLQPEDAPQPG 1016
 Db 815 TLASSATTAAPGVSTVPSQLTLLQPVTO-----LPSQVHQLLQ-----AVQSMG 863
 QY 1017 SSPRGKRSRSPAPPADKEAFAEAQKLPDPPCWTSGLPFPVPPPREVIKASHAPDPSAFS 1076
 Db 864 I-----PANL-GQAAEVLSSGD--VLYQGFPLPPOY-----FGDSN 899
 QY 1077 YAPPGHPLPLGHDHTARPVL--PRPTINPPPLISSAKHPSV---LERQIGAISOGMSV 1131

	900	IAPSNVASVCIHST---VLXPPMPTEVLATPGVFPTVPQYVESNLLVPMGGV--GGQV	954
Db			
Qy	1132	QLHVYPYSEHAKAPGVPTMGLEPLPMDPKLAPFSGVKQEQLSPRGQAGPPESLCV----	1186
Db			
Db	955	QVSQFGSLAQAPTSSQAQV-----LESTQGV-----SQVAPEAVPAQOQA	998
Qy	1187	--PTAQSASVLGRTALGSVPGGSIITKGIPSTRVPSDSAIIVRGSITHGTADLVLYKGTIT	1244
Db			
Db	999	TQPTFTTLASSV--DSAHDVASG-MSDG--NENVPSSG-----RHEGRTT	1038
Qy	1245	RIIGEDSPRLDRGEDSLPKGHVIYEGKKGHVLSYEGGMSVTQCSKE-----	1292
Db			
Db	1039	KBYRKVSRSRHEKTSRPKLRLILVSNKGD-----RVVECOLETNRKMVTFKFD	1090
Qy	1293	-DGRSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPPERHSPHHLKEQHII	1351
Db			
Db	1091	LDGN-----PEEATIMVNDFI-----LAIERESFVDQVREII---EKADEMLSEDVSV	1138
Qy	1352	RGSITQGISYVEAOEDYLRRERAKLLKREGTPPPPSPSRDLTEAYKTQALGPLKLKPAH	1411
Db			
Db	1139	EPECDOQL--ESLOCKDDYGFSGSKLEGFEKQIPASSM-----FOQIGIPT	1184
Qy	1412	EGLVATVKEAGRS--IHIEPREELRHTEPLFLAPRLKEGSIITQGTPKYDTGCASITGSK	1469
Db			
Db	1185	SSLTVVHSAGRRIIVSPVPESRLRESKVFP-----SEIT-----DTVAASTAQS	1229
Qy	1470	-----KHDEVSL-----IGSGPRTFPPVHPDLWDVADARALE	1500
Db			
Db	1230	PGMNLSHSASLSLQQAFSLBRAQMTEGPNTAPPNFSTHTPTFPVVVP-----	1278
Qy	1501	RACVEESLKRPCTASSSGGSIARGAP-----VIVPELGKRQOSPL--TYEDHGAPFA	1551
Db			
Db	1279	--FLSSIAGVPTTAATAATPVATSSPPNDISITSIOSEVTVTHEGIAGVATSTGVVTS	1335
Qy	1552	GHI.PRGSPVTWREPTPLQSGSI.SSKASOQRKLTSTPREIA---KSPHETVVEHHHPHI	1608
Db			
Db	1336	GGL-----PIPPVSESPLVSVSS-----ITIPAVUSITSTSPLSQVPTSTSEIV	1381
Qy	1609	SPYEHLRLVGSDVLYRSHIPLAFDPPTSIPRGIPLDAAAAYLPRLHAPNPITYPHYLPY	1668
Db			
Db	1382	-----VSSITALYPS-----VTVSATSAGSGTATPGPK-----PPA	1413
Qy	1669	LI-----RGYPDAALENRQ-----TIINDYTQQOMH	1696
Db			
Db	1414	VSSQQAAGSTTVGATLTSVSTTTSFFSTASQLSIQLSSSTSTPTLAEATVVVSAHSLDKTS	1473
Qy	1697	HNTATAMA-----QPADMLRGLSPR-----ESSIALNYAAGPRGIID	1733
Db			
Db	1474	HSSTGLAFSLAPSSSSSPCAGVSVYSOPGGLHPLVIPSVIATSTPILPOAAGPTSTPL	1533
Qy	1734	LSQVPHLPVLVPPPTPGTPTAMDLRAYLP.TAPOFSSRSHSSSPLSPGP--THLT.KTKTTS	1792
Db			
Db	1534	LPQVPSITPLLQP.VANVEAV--..QOTLIHSOPQ-----ALLPNQPHTHCP-----	1576
Qy	1793	SSERRDRDRDRDREREKSILTSITTYEHAPIWRPTEQSGSGSGSGSGSGSGSSRPDA	1852
Db			
Db	1577	-----EVSDTQPKAPGIDDIKTLEE-----KLRSJFSBHSSGA-----	1611
Qy	1853	SHSHAHQHSPISPRTODALQORPSVLHNTGMKGI-ITAVEPSKPTVLRSTSTSSVRPAA	1911
Db			
Db	1612	-----QHASVSLETS-----LVIESITVPGIPTTAVAPSK--LLTSTT-----S	1648
Qy	1912	TFFPATHCPLGGTLDGYPTLMPEVLLPKCAPRVARPERPRADTGHAFTAKPPARGGLEB	1971
Db			
Db	1649	TC.LUPP.NLPLUG-----TVALFVTVVTPGVGVSTPVSTTTTSGVKP	1687
Qy	1972	ASSPSKGEPRPLVPPVSGHATTARTPAKNLAPHASPDPPAPPASADPHREKTS-Q-KP	2030
Db			
Db	1688	GTAPSXEPLTKAPVLPVGTLPAGTLPSEQL-----PPFGPSL-----TQSQP	1732
Qy	2031	FSTQBLELESLGHHGSSYGPEGVEPVSVSSPSLTHDKGLPKHLEBLDXHLEGELRPKQ	2090
Db			
Db	1733	LEDLDAQLR-----RTLSPXITVITSV-----	1755

QY	2091	PGPVKLGEAAHLPHLRPLPE---	GPSSPPLIQTAPGV-----KGQRVVTLAQHISEVI	2144
		:	:	
Db	1756	-GPVMAAPTA-ITEAGTQPKGVSVQKEG	PVLATSSGAGVFKGRFQVSVAA-----	1806
		:	:	
QY	2144	TQDYTRHHFPOOLSAPLPAPILSYSPG	ASCPVLDIRRPPSDILYLP--	2193
		:	:	
Db	1807	--DGAQKEGKNKSDAKSVHFESTSSSV	SSSSPESTLVKPEPNGITIPGISSDV	1864
		:	:	
QY	2194	ARGSPHSEGGKRSPEPNKTS-----	VLGGGEDGIE-----PVSPPPEGWTE	2233
		:	:	
Db	1865	AHKTTASEAKSDGTQPTKVG	RFOVTTTANKVGRFSVSKTEDKLTDTYK	1924
		:	:	
QY	2234	PGHSRAVYLLVRDGTQTPSRMGSKSP	QNTSQP--PAFFSKLITESNAVMVSKSQE	2292
		:	:	
Db	1925	EQAVLPAPVIPKKEKP-ELSEPSHLN---	GPSDDPEAAFLSRDVEDDGS	1979
		:	:	
QY	2293	KLANTHNEPEYNISOPGTIEIKNP	PAITGTLMTYRSQAVQ-----EHASTNMGLEA	2344
		:	:	
Db	1980	SL-----PSQNLSQLSNSFNSSYN	SDNESDIEDDLKLELRRLDKHLKEI	2032
		:	:	
QY	2345	IIRKALMGKYDQWEEESPPLSANAF	NPLNASASLPAAMPITAAAGRSDHTLT	2404
		:	:	
Db	2033	RQKHEISLTYLKLGVPP-----	AVIIPPAPLS---GRRRRPTKS-----	2073
		:	:	
QY	2405	VSRGPPSRKAKSP-APGLASGDRPP	SV-----SSVHSEGD	2438
		:	:	
Db	2074	KSRSSSLGNKSPQLSGNLSCGSAAS	VLVHPQOTLHPPGN	2112
		:	:	
RESULT 99				
ABO44381				
ID	ABO44381 standard; protein; 2382 AA.			
AC	ABO44381;			
XX				
DT	26-SEP-2003 (first entry)			
XX				
DE	Novel human protein kinase #1.			
XX				
KW	Human; kinase; enzyme; cosmetic application; nutraceutical application.			
OS	Homo sapiens.			
XX				
FN	US6541252-B1.			
XX				
PD	01-APR-2003.			
PF	14-MAY-2001; 2001US-00854856.			
XX				
PR	19-MAY-2000; 2000US-0206015P.			
XX				
PA	(LEXI-) LEXICON GENETICS INC.			
XX				
PI	Walke DW, Hilbun B, Donoho G, Turner CA;			
XX				
DR	WPI; 2003-575927/54.			
DR	N-PSDB; ACH03764.			
XX				
PT	New nucleic acid encoding novel human proteins, useful in cosmetic and nutraceutical applications.			
PS	Disclosure; Page; 11pp; English.			
CC	The invention relates to a new isolated nucleic acid encoding a novel human protein kinase. The nucleic acid is useful in cosmetic and nutraceutical applications. The present sequence represents the amino acid sequence of a novel human protein kinase. Note: The sequence data for this patent did not from part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6541252B1			
XX				
SQ	Sequence 2382 AA;			

Query Match	3.1%; Score 408; DB 7; Length 2382;		
Best Local Similarity	19.6%; Pred. No. 1.2e-13;		
Matches	513; Conservative 291; Mismatches 943; Indels 872; Gaps 114;		
QY	147 PVSPPSPHTDPELELVPPRLSKLEELIQNMDRVDREITWVQIQISKKKQOOLBEEAAK 206	QY	1017 SSPROKSRSPADKAPAEAAQKLPDPPCWTSGLPFPVPPREVIVKASPHAPDPSA 107
DB	99 PLSLPQSPISAAVQSPAPPEHRETV-----TATATQVAQPPAAAPGEQAVA 149	DB	924 I-----PANK-GQAAEVPLSSGD--VLYQGFPLRPQY-----PGDSN 959
QY	207 PREP-----EKVSPPPPIESKHSVLVQIYDENRKKAAEAHRLGLEGLQFQVELPLY 257	QY	1077 YAPPGHPLPLGLHDTARPVL--PRPTISNPPLISSAKHPSV--LERQIGAISQMSV 1131
DB	150 GRAPSTVPSSTKDRVPSQSL-----VSKKEPPPA 181	DB	960 IAPSNVASVCIHST---VLXPPMTEVLATPGYPTVVQPVVSVESNLLVPMGV--GGQV 1014
QY	258 NOPSDFRQYHENIKINQAMRKKLILYFKRNHARKQWKQFCQRYDQLMEALKKVERIE 317	QY	1132 QJHVYVSEHAKAPGVGVTMGLPLMDPKKLAPFGSKQBQLSPRGQAGPPESLGV----- 1186
DB	182 RSGSG-----GGSAKEPQBERSQQDDI-EELETAKVAGMS 215	DB	1015 QVSPGGSLAQAPTSSQAV-----LESTQGV-----SQVAPAEVAVAQPOA 1058
QY	318 NNPRRAKESKV-REYVEKQFBEIRKQ-----ELQRMQSRVQORSGLSMSARSE 369	QY	1187 ---PTQOBSVLRGTLGVSPPGSGITKGPSTRTVPDSAITYRGSITHGTPADVLVYGTIT 1244
DB	216 NDRFLKFDIEIGRSGFKTVYKGLDTEVAVACLODRK-----LTKSERQRF 265	DB	1059 TQPTTLASSV--DSAHSDVASG--NENVPSSG-----RHEGRIT 1098
QY	370 HEVSEIIDGLSEKENLEKQMLAVIPMLYDADQORIK-----FINMGLM-ADPMKVYK 424	QY	1245 RIIGEDSPSLDRGREDSLPKGHVIEGKGVLSYEGGMSVTOCSKE----- 1292
DB	266 KEAEMLKGL-QHPNIVR-----FYDSWESTVKGKICIVLTMLSGTUKTYL 313	DB	1099 KKHVRSVRSRHEKTSRPKLRIILNVSNKGD-----RVVECOLETHNRKMTFKPD 1150
QY	425 DR-----QVMNMWSQEKETFREKFMQHPK-----NFGLIA 455	QY	1293 -DCRSSGGPPHETAAPKRTYDMMEGRVGRRAISSASIEGLMGRAPPPERHSPHHLKEQHII 1351
DB	314 KRFKVMKIKVLRSMCRQILKGLQFLHRTPTIHRDLKCDNIFITGPTGVSXIGDGLG-- 371	DB	1151 LOGDN-----PEEIAITMVNDFI-----LAIERESFVDQVREII---EKADEMLSEDVSV 1198
QY	456 SFLERTVAECVL-----YYLYTKQENY-----KSLVRSVRRRGKSGQQOQQOQ 501	QY	1352 RGSITQGIPIRSYVBAQEDYLREAKLLKREGTPPPPPPPSRDLTEAYKTOALGPLKLPKH 1411
DB	372 ATLKRASFVIGTPEFMAPEYEEKYDESDVYAFGCMLEMATSEYPISECQNAQI 431	DB	1199 EPEGDQGL--ESLQKDDYFGSGQKLEGEFKQPIPASSM-----PQIGIPT 1244
QY	502 QOQOQOQOQO-----MPSRSG-----BEKDEKEKEKEKEKEKEKEKEKEDLLK 547	QY	1412 EGLVATVKEAGS--IHEIPRELHRTPELPLAPRLKEGSIQTGTPLYKVDGASTGSK 1469
DB	432 YRVTSGVPASFDKVAIBEVEKIIIEGTRONKDERYSIKDLNHAFFQETGVRVELAE 491	DB	1245 SSLTQVTHSAGRRFVSPVPSRLRESKVFP-----SEIT-----DVAASATQOS 1289
QY	548 EKTDDTSGEDNDEKAEVASKRGKRTANSQRRKGRITRSMANEANSE-EAITPQCSAELAS 606	QY	1470 ----KHDVRSI-----IGSPGRTPFPVPHPLDVNMADARALE 1500
DB	492 E-----DDGEKIAIKLWLRIEDIKLKGKYNDEALIEFSLDERDVEDVAQ---- 538	DB	1290 PGMNLSHSSASSLSQAFSELRAQWTEGTPAPNFSHTGTPFPVPP----- 1338
QY	607 MELNESRWTEBEMETAKGLLEHGRNWSAIAARMVGSKTVSQCKNFYFNKGRQNLDEL 666	QY	1501 RACYEESLKSRTGTSASSGSGIARGAP-----VIVPELGKPRQPL--TYEDHGAPFA 1551
DB	539 -BWESVYCEGDHKTMAKIDR-----VSLIK-----RKREQRQLVR 576	DB	1339 ---FLSSIAGVPTTAAATAPVATSPNDISTSVQSESVTVPTTEGIAAGVATSTGVVTS 1395
QY	667 QQHLKWEKERNARRKKKA-----PAAASEEAFFPPVVEDEMEASG 709	QY	1552 GHLPRGSPVMEPTPRLOEGSLSSKASQDRKLTSTPREIA---KSPHSTVPEHHPHI 1608
DB	577 EQEKKKKQESSLKKQVQESSASQGTGKOLPSASTGIPASTTSASVSTQVEPEPEEA-- 634	DB	1396 GGL-----PIPPVSESPVLSVSS-----ITIPAVSISTTSPLSQVSTSTSEIV 1441
QY	710 VSGNEEMVVEEALHASGNEVPRGCSGPATVN--NSSDTEIPSPHTEAAKDTQONGP 767	QY	1609 SPYEHLLRGVGVLYRSHIPLAFDPTSIPIRGIPDLDAAYLYLPHLAPNPTYPHLYPPY 1668
DB	635 -DOHQQLYQOQPSISVLSDGTVDGSGSVFTESRVSSQTVSYGSGH-EQAHSTGTVP 692	DB	1442 ----VSSTALYPS-----VTVSATSAGSGSTATGPK-----PPA 1473
QY	768 KPPATLGADGPPG--PPT-----PPR 787	QY	1669 LI-----RGYPDTAALENRQ-----TIINDYITQQMH 1696
DB	693 HIPSTVQAGSQHGVPPSPSAQOGQSGQSSSLTGVSSSQIOHPQOQOQIQTAPPQ 752	DB	1474 VVSQAAGSTTVGATLTSVSTTSPSTASQSLQSSSTSTPTLAETVVSASHSLDKTS 1533
QY	788 RISRARIET-PASEATGAPTPPAPSPAPPPVVPVKEKEEETAAAPVE--EGEOK 844	QY	1697 HNTATAMA-----QADMLRGLSPR-----ESSIALNYAAGRGID 1733
DB	753 QTVQVLSQTSSTSEATTA-----QVSPQAPQVLPQVSAQKLPVSPVPTIQESQPI 807	DB	1534 HSTTGLAPSLAPSSSSSPGAGVSYSIQPGGLHPLVIPSVIASPTLPQAAGPTSTPL 1593
QY	845 PPAABE--LAVDTGKAEPVKSECTEABEGPAKGDAEAAEATARGALKAEKKGSGR 902	QY	1734 LSGVPHLPVLVPTPGTPATAMDRLAYLPTAPQSSRSHSSSPSLSPGPG--THLTUKPTTS 1792
DB	808 PVATQSVVPHSGAFLPV-----GQPL 831	DB	1594 LPQVPSIPFLVQPVANPAV---QQTLIHSQGP-----ALLNQPHTHCP----- 1636
QY	903 ATPAKSSGAPQSDSATSADVEDBAEGDKNRLSPRSLTPTGDPNANSPKPLD 962	QY	1793 SSERERDRDRDREREKSKILTSTTTVEHAPIWRPGTEQSSGSGSGSGGSGSSRPA 1852
DB	832 PTP-----LPQVPSQIPISTPHVSTAQT-----FSSLPIITMA-----AGITQPLL 874	DB	1637 -----EVDSDTQPKAPGIDDIKTLEE-----KLRSFSEHSSGA----- 1671
QY	963 LKQLKORAAIP-----PQVTKVHEPPREDAATPKAPPAPPPQNLOPESDAPQOQ 1016	QY	1853 SHSHAHQHSPISPRITQDALQORPSVLHNTGMKI-ITAVEPSKPTVLRSTSTSSVPRAA 1911
DB	875 TLASSATTAAIPGVSTVPSQLPTLQPVTVQ-----LPQVHPQLQPP-----AVQSMG 923	DB	1672 -----QHASVSLETS-----LVIESTVPTGIPITAVAPSK--LLTSTT-----S 1708
		QY	1912 TFPPTHCPGLGGTLDGVYPTLMEPVLLPKAEAPVARPERPRADTGHAFKAPPARSGLEP 1971
		DB	1709 TCLPPTNLPLG-----TVALPVTVVTPGVSTVSTTSGVKP 1747
		QY	1972 ASSPSKSGSEPRFLVPPVPSGHATITARTPAKNLAPHASPPPPASADPHREKTOS-KP 2030

Db 1748 GTAPSKPLTKAPVLPVGTLPAGTLFSEQL-----PPFPSPSL-----TQSQOP 1792
Qy 2031 FSIQELRLSLGVHGGSSYPGEVPSVPSSPSTHDKGLPKHLELDKSHLGEGLRPQ 2090
Db 1793 LEDLDAQLR-----RTLSPEXITVTSV----- 1815
Qy 2091 PGPKVLGGEAAHLPLRLPE---SOPSSPPLLQTAPGV---KGHVVVTLAQHISEVI 2143
Db 1816 -GPVMAAPTA-ITEAGTQPKGVSVQVKEGVLATSSGAGVFKWGFQVSVAA----- 1866
Qy 2144 TDYTRHHPQOLSAPLPAPLYSPFGASCVPVLDLRRPESDLYLPPP-----DHGAP 2193
Db 1867 --DGAQKEGKNKSEDAKSVHFESSTSESSVLSSSPSESTLVKPEPNGITIPGISDVPS 1924
Qy 2194 ARGSPHSEGGKRPENKTS-----VLGGGEDGIE-----PVSPPGEMTE 2233
Db 1925 AHKTTASEAKSDGTQPKTKGRFOVTTANKVGRFSVSKTEDKITDTKKGVPVASFPM 1984
Qy 2234 PGHSRAVYLLYRDGEQTEPFRMGSKSPGNTSQP--PAFFSKLTESNAWKSKEINK 2292
Db 1985 EQAVLPAVIPKKEP-ELSEPHLN---GPSSDPEAFILSRDVDGSGSPHQLSK 2039
Qy 2293 KLTHNRNEPEYNISQGTETFMNPAITGTLMTYRSQAVQ-----EHASTNMGLEA 2344
Db 2040 SL-----PSQNLSQLNSFNSSYMSDNEIDEDLKLELRLRLDKHLKEIQDLQS 2092
Qy 2345 IIRKALMGKYDOWEESPPLSANAFNPLNASASLPAAMPITADGRSDHTLTSPGGGKAK 2404
Db 2093 RQHEIESLYTLKGVPP-----AVIIPPAAPLJ-----GRRRRTKS-----KGS 2133
Qy 2405 VSGRPSRKAKSP-AFCLASGDRPPSV---GSVHSEGD 2438
Db 2134 KSKRSSSLGNKSPQLSGNLSGQSAASVLPHPQTLHPGN 2172

RESULT 100
ID AAM41059
XX AAM41059 standard; protein; 1565 AA.
AC AAM41059;
XT
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5990.
XX
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
FN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60215.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
PS Example 2; SEQ ID NO 5990; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1565 AA;
Query Match 3.1%; Score 406.5; DB 4; Length 1565;
Best Local Similarity 20.8%; Pred. No. 8.7e-14;
Matches 368; Conservative 188; Mismatches 612; Indels 597; Gaps 82;
Qy 355 GORGSLMSAARSEHVESEIIDG---LSQENLEKQMRQLAVIPPLMDAQOQIRKI 410
Db 10 GARGSGWG-SRSQAPYGTGLGAVSGGEQVLLHEEAGDSGFVLSRLGPSLRDKLEMEELM 68
Qy 411 NMNGLMADPMKVYKDRQVMNMWSEKETTFREKFMQHPKNGFLIASFLERKTVACVLYY 470
Db 69 LQETLLGTQSYMDASLISLIEDFGSLGEVEMSLPDPDFSPFPLETS----- 119
Qy 471 YLTKNENYKSLVRRSRRRGKSSQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 510
Db 120 --SPKLPSMRP--PRSRFRWQSPPPQQRSDGEEVEVAFSGQILAGELNCVSSIPDF 175
Qy 511 PMPRSSQBEKE 545
Db 176 PMHLACPEED-----KATAAEMAVPAAGDSSISSELSELVRAHPYCLPNTLHSLASLE 230
Qy 546 LKEKTDD-----TSGED----- 557
Db 231 LQEPDDLTLPQGCVVLEIVQAAATAGDDLEIPVVVQVSPGPRPVLLDSSLETSSALQL 290
Qy 558 -----NDEKEAVASKGRKTANSQGRKGRITRSMANEANSEAI----- 596
Db 291 LMPTLESETEAAVPKVTLCSEKGL-----SLNSEKLDLSACLKPREVPEPV 339
Qy 597 -----TPQOQSAELASMEINSSRWTEEMETA-KGGLLEHGRNWSAIAWMVGSQK 650
Db 340 PKFPQNPANAAAGSORARKGRKKKKEQPAACVEGYARRLRSSRSGQSTVGTEVTSQ-- 397
Qy 651 NFYFNKKRQNLDBILQOHKLKMEKERNARRKKKAPAAASEEAAFPVVEDEMEASGV 710
Db 398 -----VDNLQPOQELQESGPLQKQKPRAWARAA-----ALENSSP 438
Qy 711 SGNEEMVVEAEALHASGNEVPRGECSPATV--NNSSDT--ESIPSPHTEAAKTGQNGPK 768
Db 439 KNLE-----RSAGQSSPAKE--GPLDLVPLADTIQTNPIPTHLSLVDSQAQSPM 486
Qy 769 PPATLGADGPPGP-----PTPP-----RRTSRAPTEPTASEATGAPTPPPAPPSFA 817
Db 487 PVDSEADPTAVGPVLGAPVPVDFGLVDLASTSSSELVEPLPAEPVL---INPVLADSAV 543

This Page Blank (uspto)